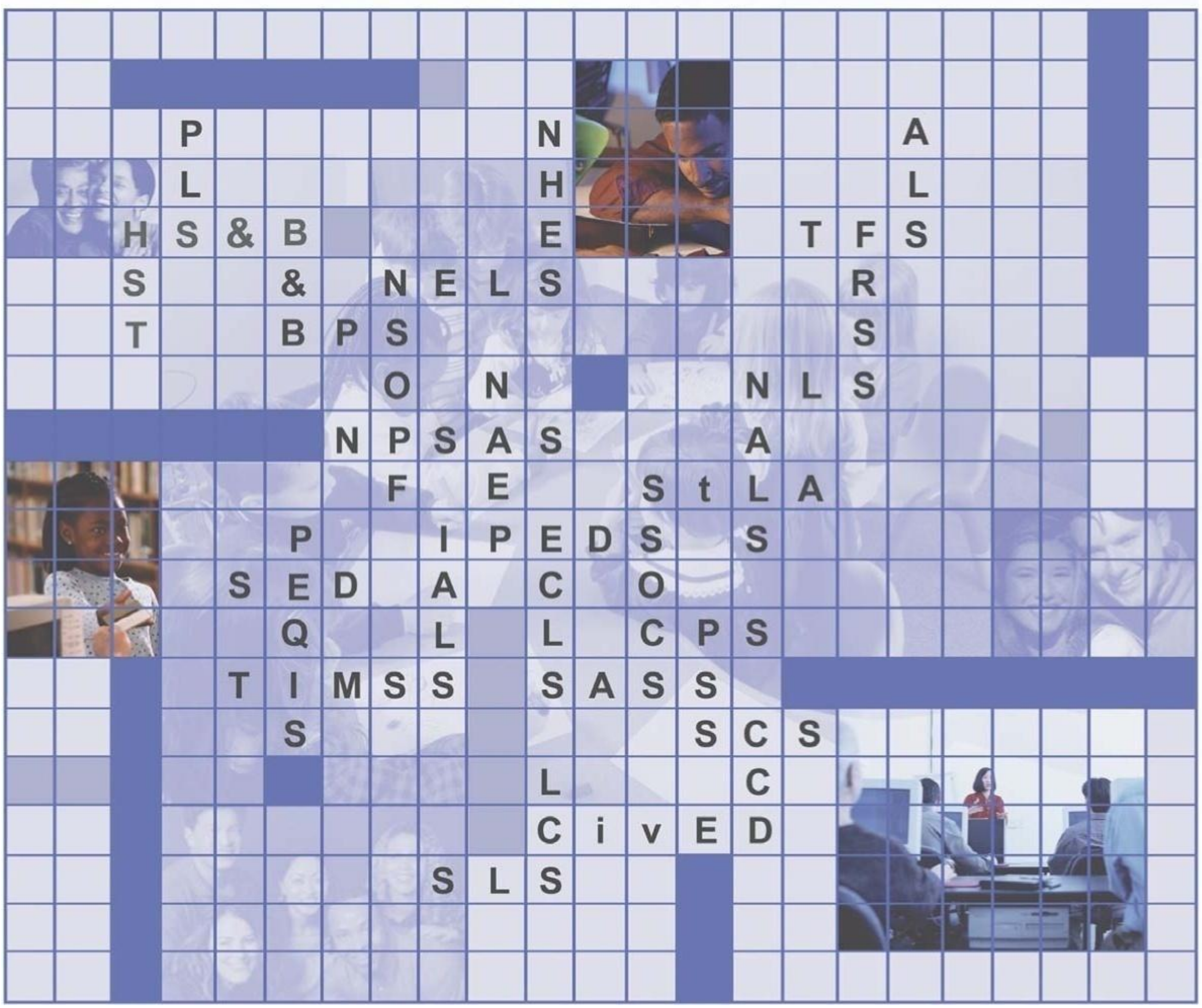
NCES Handbook of Survey Methods



NCES 2011-609

U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

NCES Handbook of Survey Methods

June 2011

Shelley Burns
Editor
National Center for Education Statistics

Xiaolei Wang
Editor
American Institutes for Research

Alexandra Henning
Editor
Quality Information Partners, Inc

NCES 2011-609
U.S. DEPARTMENT OF EDUCATION

NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Statistics
Jack Buckley
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

> NCES, IES, U.S. Department of Education
> 1990 K Street NW
> Washington, DC 20006-5651

JUNE 2011

The NCES Home Page address is http://nces.ed.gov.
The NCES Publications and Products address is http://nces.ed.gov/pubsearch.

This report was prepared in part under Contract No. ED-05-CO-0044 with American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation
Burns, S., Wang, X. and Henning, A. (Eds.) (2011). *NCES Handbook of Survey Methods* (NCES 2011-609). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

For ordering information on this report, write to

> ED Pubs, U.S. Department of Education
> P.O. Box 22207
> Alexandria, VA 22304

or call toll free 1-877-4-ED-Pubs or order online at http://www.edpubs.gov.

Content Contact
Marilyn Seastrom
(202) 502-7303
Marilyn.Seastrom@ed.gov

Acknowledgments

This second edition of the *NCES Handbook of Survey Methods* would not have been possible without the efforts of many people. Sincere thanks to all the NCES project officers on the surveys and studies documented in the Handbook—who answered questions, provided unpublished materials, and reviewed the chapters about their surveys.

Contributing Authors

Xiaolei Wang	American Institutes for Research
Alexandra Henning	Quality Information Partners, Inc.
Weiwei Cui	National Institute of Statistical Sciences
Francis Huang	Formerly of American Institutes for Research
Susan Armstrong	American Institutes for Research
Kyo Sin Kang	American Institutes for Research
Jinae Boyer	American Institutes for Research
Simone Robers	American Institutes for Research

Contents

LIBRARY SURVEYS

POSTSECONDARY AND ADULT EDUCATION SURVEYS

EDUCATIONAL ASSESSMENT SURVEYS

HOUSEHOLD SURVEYS

SMALL SPECIAL-PURPOSE NCES SURVEYS

APPENDICES

List of Acronyms

2PL	Two-parameter logistic [model]
3PL	Three-parameter logistic [model]
ACA	American Correctional Association
ACER	Australian Council for Educational Research
AE	Adult Education [survey]
AELL	Adult Education and Lifelong Learning Survey
AEWR	Adult Education for Work-Related Reasons [survey]
ACG	Academic Competitiveness Grant
AI/AN	American Indian and Alaska Native
ALA-ORS	American Library Association's Office of Research and Statistics
ALL	Adult Literacy and Lifeskills Survey
ALS	Academic Libraries Survey
ALSA	Adult Literacy Supplement Assessment
AM	analytical software
AP	Advanced Placement
AQS	Attachment Q-Sort
ASCII	American Standard Code for Information Interchange (character-encoding scheme used for representing data as text)
ASPA	Before- and After-School Programs and Activities Survey
B&B	Baccalaureate and Beyond Longitudinal Study
BHR	Balanced half-sample replication [method of variance estimation] (also referred to as BRR)
BHS	Balanced half-sample
BIB	Balanced Incomplete Block Spiraling (a type of matrix sampling)
BIE	Bureau of Indian Education. In 2006, the Office of the Assistant Secretary for the Department of the Interior established the Bureau of Indian Education (BIE). This organization is responsible for prekindergarten, elementary, secondary, and postsecondary schools and educational activities of the Bureau of Indian Affairs (BIA). Prior to 2006, these schools were referred to as BIA schools.
BILOG	computer software used in scaling
BJS	Bureau of Justice Statistics, U.S. Department of Justice
BLS	Bureau of Labor Statistics, U.S. Department of Labor
BPS	Beginning Postsecondary Students Longitudinal Study
BRR	Balanced Repeated Replication [method of variance estimation] (also referred to as BHR)
BSF-R	Bayley Short Form-Research edition [scale]
BSID	Bayley Scales for Infant Development
BYI	Base Year Ineligible [study]
C	Completions [survey component]
CA	Chief Administrator
CACE	Computer-Assisted Coding and Editing
CADE	Computer-Assisted Data Entry

CAPI	Computer-Assisted Personal Interview
CATI	Computer-Assisted Telephone Interview
CCD	Common Core of Data
CCO	Child Care Observation
CCP	Child Care Provider
CHAID	Chi-squared Automatic Interaction Detector
CI	Civic Involvement [survey]
CIP	Classification of Instructional Programs
CN	Consolidated Form [survey component]
CN-F	Consolidated Form, “finance” part [survey component]
COSLA	Chief Officers of State Library Agencies
CSSO	Chief State School Officer
CPS	Current Population Survey
CSSC	Classification of Secondary School Courses
CSV	Comma-separated values (file format)
DAS	Data Analysis System
DIF	Differential Item Functioning
DoDEA	Department of Defense Education Activity [schools]
DOL	U.S. Department of Labor
DPC	Data Processing Center
DRF	Doctorate Records File
E12	12-Month Enrollment [survey component]
EAP	Employees by Assigned Position
ECB	Electronic Codebook
ECE	Early Childhood Education [survey]
ECEP	Early Care and Education Provider
ECLS	Early Childhood Longitudinal Study
ECLS-B	Early Childhood Longitudinal Study, Birth Cohort
ECLS-K	Early Childhood Longitudinal Study, Kindergarten Cohort
ECPP	Early Childhood Program Participation [survey]
EDAT	Education Data Analysis Tool
ED	Enumeration District
ED Pubs	U.S. Department of Education Publications Center
EDEN	Education Data Exchange Network
EDI	Electronic Data Interchange
EEO-6	Higher Education Staff Information [report]
EEOC	U.S. Equal Employment Opportunity Commission
EF	Fall Enrollment
EFC	Expected Family Contribution
ELS: 2002	Education Longitudinal Study of 2002
EP	Fall Enrollment in Occupationally Specific Programs [survey component]
ERIC	Education Resources Information Center
ESL	English as a Second Language

ETS	Educational Testing Service
F	Finance [survey component]
FAN	Fluency Addition to NAAL
FASB	Financial Accounting Standards Board
FICE	Federal Interagency Committee on Education
FIMS	First International Mathematics Study
FIPS	Federal Information Processing Standards
FISS	First International Science Study
FRSS	Fast Response Survey System
FSES	Followback Study of Excluded Students
FTB	First-time beginning/beginners
FTE	Full-time Equivalency
FY	Fiscal year
GASB	Governmental Accounting Standards Board
GED	General Equivalency Diploma
GEM	Generalized Exponential Model
GPA	Grade Point Average
GPCM	Generalized Partial Credit Model
GPO	Government Printing Office
GR200	Graduation Rates 200
GRL	Graded response logistic [model]
GRS	Graduation Rate Survey
GSS	General Social Survey
GVF	Generalized Variance Function
HBCU	Historically Black Colleges and Universities
HEGIS	Higher Education General Information System
HEOS	Higher Education Opportunity Act
HHL	Household and Library Use [survey]
HLM	Hierarchical Linear Modeling
HR	Human Resources [survey component]
HS&B	High School and Beyond
HSES	High School Effectiveness Study
HSI	Hispanic-Serving Institutions
HST	High School Transcript [studies]
IAEP	International Assessment of Educational Progress
IALS	International Adult Literacy Survey
IB	International Baccalaureate
IC	Institutional Characteristics [survey component]
ICS	Integrated Control System
ICT	Information and communication technology
IDE	International Data Explorer
IDEALS	Input and Data Editing for Academic Library Statistics
IEA	International Association for the Evaluation of Educational Achievement

IEP	Individualized Education Program
IES	Institute of Education Sciences
IHE	Institution of Higher Education
IMS	Integrated Management System
IPEDS	Integrated Postsecondary Education Data System
IRT	Item Response Theory
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
ISIC	International Standard Industrial Classification
ISR	Interview status recode
JRR	Jackknife Repeated Replication
LEA	Local Education Agency
LEP	Limited English Proficiency
MDR	Market Data Retrieval, a Dun & Bradstreet (D&B) Company
MEP	Migrant Education Program
MIL	Missing Information Letter
MOL	Mathematics Online [study]
MOS	Measure of size
MML	Marginal maximum mikelihood
MSA	Metropolitan Statistical Area
NAAL	National Assessment of Adult Literacy
NAEP	National Assessment of Educational Progress
NAGB	National Assessment Governing Board
NALS	National Adult Literacy Survey
NCATS	Nursing Child Assessment Teaching Scale
NCEA	National Catholic Educational Association
NCES	National Center for Education Statistics
NCLIS	U.S. National Commission on Libraries and Information Science
NCOA	National Change of Address (data base)
NCR	National Research Council
NCVS	National Crime Victimization Survey
NDE	NAEP Data Explorer
NELS:88	National Education Longitudinal Study of 1988
NHES	National Household Education Survey
NIES	National Indian Education Study
NIH	National Institutes of Health
NLS:72	National Longitudinal Study of the High School Class of 1972
NORC	National Opinion Research Center
NPEFS	National Public Education Financial Survey
NPM	National Project Manager
NPS	New Participant Supplement
NPSAS	National Postsecondary Student Aid Study

NPSQ	New Participant Supplement Questionnaire
NRC	National Research Council (chapters 17 and 20); National Research Coordinator (chapter 25)
NSF	National Science Foundation
NSOPF	National Study of Postsecondary Faculty
NTID	National Technical Institute for the Deaf
OBE	Office of Business Economics
OCR	[U.S. Department of Education's] Office for Civil Rights
OECD	Organization for Economic Cooperation and Development
OLDS	Oral Language Development Scale
OMB	Office of Management and Budget
OPAC	Online Public Access Catalog
OPEID	An 8-digit school identification code developed by OPE
OREALC	Regional Office of Education for Latin America and the Caribbean [UNESCO] [Acronym is based on foreign phrase]
OPE	Office of Postsecondary Education
PC CARP	Analytical software
PEPS	Postsecondary Education Participants System
PEQIS	Postsecondary Education Quick Information System
PETS	Postsecondary Education Transcript Study (chapter 6); Postsecondary Education Telephone System (chapter 12)
PFI	Parent and Family Involvement in Education [survey]
PIAAC	Program for the International Assessment of Adult Competencies
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
PLS	Public Libraries Survey
PPA	Program Participation Agreement
PPS	Probability proportional to size (a sampling method)
PROC IMPUTE	A regression-based imputation method
PSS	Private School Universe Survey
PSU	Primary Sampling Unit
QED	Quality Education Data, Inc.
QCM	Quality Control Monitor
R	analytical software
RA	Research assistantship
RCS	Receipt Control System
RDD	Random digit dialing (telephone survey technique)
ROTC	Reserve Officers Training Corps
RTD	Registered time-to-degree
RTI	Research Triangle Institute
S	Fall Staff [survey component]
SA	Survey administrator (chapter 9); Salaries [survey component] (chapter 12)
SAA	Survey administrator assistant
SAAL	State Assessment of Adult Literacy

SAQ	Self-administered questionnaire
SAS	analytical software
SASS	Schools and Staffing Survey
SAVD	School Associated Violent Death Study
SCS	School Crime Supplement
SD	Students with Disabilities
SDR	Students with Disabilities
SEA	State Education Agency
SED	Survey of Earned Doctorates
SEOG	Supplemental Educational Opportunity Grants
SES	Socioeconomic status
SFA	Student Financial Aid
SHR	Supplementary Homicide Reports
SIF	School Information Form
SIMS	Second International Mathematics Study
SIPP	Survey of Income and Program Participation
SISS	Second International Science Study
SLS	School Library Survey
SMART	Science and Mathematics Access to Retain Talent [Grant]
S-Plus	analytical software
SPSS	analytical software
SQL	Structured Query Language
SR	School Readiness [survey]
SRIF	Student Record Information Form
SS&D	School Safety and Discipline [survey]
SSOCS	School Survey on Crime and Safety
SSP	Stratification Search [analytical] Program
SST	Secondary School Taxonomy
STATA	analytical software
STEM	Science, Technology, Engineering, and Mathematics
SQL	Structured Query Language
SUDAAN	analytical software
SUREG	a command in computer program STATA
TAS-45	Toddler Attachment Sort-45
TBA	Technology-Based Assessment [project]
TFS	Teacher Follow-up Survey
TIGER	Topologically Integrated Geographical Encoding and Referencing
TIMSS	Trends in International Mathematics and Science Study
TRE	Technology-Rich Environments [study]
TSA	Trial state assessment
TTD	Total time-to-degree
TUDA	Trial Urban District Assessment
UCR	Uniform Crime Reporting

UH	Une Heure (one-hour) [booklet]
UIS	Institute for Statistics
UNESCO	United Nations Educational, Scientific, and Cultural Organ
USDA	U. S. Department of Agriculture
USU	Ultimate Sampling Unit
WesVar	analytical software
WISQARS	Web-based Injury Statistics Query and Reporting System
WOL	Writing Online [study]
YALS	Young Adult Literacy Survey
YCI	Youth Civic Involvement [survey]
YRBSS	Youth Risk Behavior Surveillance System

Introduction

Since its inception, the National Center for Education Statistics (NCES) has been committed to the practice of documenting its statistical methods for its customers and of seeking to avoid misinterpretation of its published data. The reason for this policy is to assure customers that proper statistical standards and techniques have been observed, to guide them in the appropriate use of information from NCES, and to make them aware of the known limitations of NCES data. This second edition of the *NCES Handbook of Survey Methods* continues this commitment by presenting descriptions of how each survey program in NCES obtains and prepares the data it publishes.

NCES statistics are used for many purposes. This handbook aims to provide users of NCES data with the most current information necessary to evaluate the suitability of the statistics for their needs, with a focus on the methodologies for survey design, data collection, and data processing. It is intended to be used as a companion report to *Programs and Plans of the National Center for Education Statistics,* which provides a summary description of the type of data collected by each program at the Center.

LAYOUT OF HANDBOOK CHAPTERS

- Overview
- Uses of Data
- Key Concepts
- Survey Design
- Data Quality and Comparability
- Contact Information
- Methodology and Evaluation Reports

NCES's Role and Organization

Among federal agencies collecting and issuing statistics, NCES is the primary federal entity for collecting and analyzing data related to education. The Center's data serve the needs of Congress, other federal agencies, national education associations, academic education researchers, public and private education institutions, tutors, education administration bodies, business, and the general public. NCES is a component of the Institute of Education Sciences (IES) within the U.S. Department of Education.

Within NCES, the Statistical Standards Program, under the direction of the NCES Chief Statistician, provides expertise in statistical standards and methodology, technology, and customer service activities across subject-matter lines. The specific survey programs of NCES, however, have developed around subject-matter areas. As a result, except for the Statistical Standards Program, NCES is organized according to these subject-matter areas, with each survey program falling under one of the following four NCES divisions:

- Assessment
- Early Childhood, International, and Crosscutting Studies
- Elementary/Secondary and Libraries Studies
- Postsecondary Studies

Organization of the Handbook

The handbook contains 30 chapters. Chapters 1 to 27 each focus on one of the 27 major NCES survey programs. To facilitate locating similar information for the various programs, the information in each of these chapters is presented in a uniform format with the following standard sections and headings:

1. *Overview.* This section includes a description of the purpose of the survey, the type of information collected in the survey, and the periodicity of the survey.

2. *Uses of Data.* This section summarizes the range of issues addressed by the data collected in the survey.

3. *Key Concepts.* This section provides the definitions of a few important concepts specific to the survey.

4. *Survey Design.* This section describes the target population, sample design, data collection and processing procedures, estimation methods, and future plans for the survey. Note that the handbook does not include a list of the data elements collected by each survey. That information can be found in the survey questionnaires, electronic codebooks, data analysis systems, or technical documentation, many available through the NCES website (http://nces.ed.gov). However, some general remarks about the data collected can be made here:

   - All race/ethnicity data are collected according to Office of Management and Budget (OMB) standards. For all surveys, data on individuals can be disaggregated by "Black," "White," "Hispanic", and "Other"; for some surveys, data can also be disaggregated by "Asian/Pacific Islander" And "American Native or Alaska Native".

   - All data on individuals can be disaggregated by sex.

   - All elementary/secondary student-level data collections include information on limited English proficiency and student disability.

   - School-level data collections include information on programs and services offered.

5. *Data Quality and Comparability.* This section describes the appropriate method to use for estimating sampling error for sample surveys and presents important findings related to different types of nonsampling error (such as coverage error, unit and item nonresponse error, and measurement error). In addition, this section provides summary descriptions of recent design and/or questionnaire changes as well as information on the comparability of similar data collected in other studies.

6. *Contact Information.* This section lists the name of the main contact person for each survey along with a telephone number, e-mail address, and mailing address. Note that at NCES, telephone numbers are assigned according to survey program; staff members leaving one survey program for another have to change telephone numbers.

   To find out the current number for a particular staff member, see the NCES Staff Directory (http://nces.ed.gov/ncestaff). To find out the current contacts for a particular survey program, please check the program's website. (NCES survey website addresses are listed in appendix D.)

7. *Methodology and Evaluation Reports.* This section lists the primary recent methodological reports for the survey. Use the NCES number provided to find a particular report through the NCES Electronic Catalog (http://nces.ed.gov/pubsearch). Each NCES survey website also contains a list of that survey's publications.

Note that some of the chapters include cautions to data users. The cautions usually appear in Section 5: Data Quality and Comparability. For example, in chapter 12, section 5, caution is urged when comparing institutions for which data have been imputed for the Integrated Postsecondary Education Data System (IPEDS), since these data are intended for computing national totals and not intended to be an accurate portrayal of an institution's data.

The first 27 chapters are organized under the following subject-matter rubrics:

- Early Childhood Education Survey

  Chapter 1: Early Childhood Longitudinal Study (ECLS)

- Elementary and Secondary Education Surveys

  Chapter 2: Common Core of Data (CCD)

  Chapter 3: Private School Universe Survey (PSS)

Chapter 4: Schools and Staffing Survey (SASS)

Chapter 5: SASS Teacher Follow-up Survey (TFS)

Chapter 6: National Longitudinal Study of the High School Class of 1972 (NLS:72)

Chapter 7: High School and Beyond (HS&B) Longitudinal Study

Chapter 8: National Education Longitudinal Study of 1988 (NELS:88)

Chapter 9: Education Longitudinal Study of 2002 (ELS:2002)

- Library Surveys

Chapter 10: SASS School Library Survey (SLS)

Chapter 11: Academic Libraries Survey (ALS)

- Postsecondary and Adult Education Surveys

Chapter 12: Integrated Postsecondary Education Data System (IPEDS)

Chapter 13: National Study of Postsecondary Faculty (NSOPF)

Chapter 14: National Postsecondary Student Aid Study (NPSAS)

Chapter 15: Beginning Postsecondary Students (BPS) Longitudinal Study

Chapter 16: Baccalaureate and Beyond (B&B) Longitudinal Study

Chapter 17: Survey of Earned Doctorates (SED)

- Educational Assessment Surveys

Chapter 18: National Assessment of Educational Progress (NAEP)

Chapter 19: National Adult Literacy Survey (NALS)

Chapter 20: National Assessment of Adult Literacy (NAAL)

Chapter 21: Trends in International Mathematics and Science Study (TIMSS)

Chapter 22: Program for International Student Assessment (PISA)

Chapter 23: International Adult Literacy Survey (IALS)

Chapter 24: Adult Literacy and Lifeskills (ALL)

Chapter 25: Progress in International Reading Literacy Study (PIRLS)

- Household Surveys

Chapter 26: National Household Education Surveys (NHES) Program

Chapter 27: Current Population Survey (CPS)—October Supplement

Chapters 28 through 30 cover multiple surveys or survey systems. The format is similar to that for chapters 1 to 27, but is somewhat abbreviated to allow adequate coverage of multiple surveys within each chapter.

- Small Special-Purpose NCES Surveys

Chapter 28: Crime and Safety Surveys: School Crime Supplement (SCS) and School Survey on Crime and Safety (SSOCS)

Chapter 29: High School Transcript (HST) Studies

Chapter 30: Quick Response Information System

Details of three surveys are not available at the time of publication, and thus not included in this version of Handbook. The High School Longitudinal Survey (HSLS:09) is a nationally representative, longitudinal study of more than 21,000 ninth graders in 940 schools who will be followed through their secondary and postsecondary years. The study focuses on understanding students' trajectories from the beginning of high school into postsecondary education, the workforce, and beyond. What students decide to pursue when, why, and how are crucial questions for HSLS:09, especially, but not solely, in regards to science, technology, engineering, and math (STEM) courses, majors, and careers. This study includes a student assessment in algebraic skills, reasoning, and problem solving, and surveys of students, their parents, math and science teachers, school administrators, as well as school counselors. The first wave of data collection for HSLS:09 began

in the fall of 2009. The next data collection will occur in the spring of 2012.

The Beginning Teacher Longitudinal Study (BTLS) follows a cohort of beginning public school teachers, who were initially interviewed as part of the 2007–08 Schools and Staffing Survey, over a decade as they continue in pre-K-12 teaching or change careers. In the 2007-08 school year, approximately 2,000 beginning public school teachers responded to a variety of questions about themselves, their schools, their preparation, struggles and future plans. The second year of data collection was 2008–09 and was included in the Teacher Follow-up Survey (TFS). Of the two questionnaires for teachers who began teaching in 2007, one was for teachers who left teaching since the previous SASS and the other for those who were teaching either in the same school as the previous year or in a different school. The topics for the Current Teacher questionnaire included teaching status and assignments, ratings of various aspects of teaching, reasons for moving to a new school, information on having had a mentor teacher in the previous year, and earnings. The topics for the Former Teacher questionnaire included employment status, ratings of various aspects of teaching and their current jobs, information on decisions to leave teaching, whether they had applied for a teaching position, and information on having had a mentor teacher in the previous year.The third year of data collection covered the 2009–10 school year. Current teachers were asked questions regarding teaching status and assignments, their opinions of various aspects of teaching, reasons for moving to a new school, reasons for returning to teaching (if they left after the 07–08 school year but returned for the 2009–10 school year), earnings, and information on having and serving as a mentor. Fomer teachers were surveyed on current employment status, their opinions on various aspects of teaching and their current jobs, information on decisions to leave teaching (if they left after the 08–09 school year), and whether they had applied for a new teaching position.

The Program for the International Assessment of Adult Competencies (PIAAC) is a cyclical, large-scale, direct household assessment under the auspices of the Organization for Economic Cooperation and Development (OECD). The assessment will be first administered in 2011 to approximately 5,000 individuals between the ages of 16 and 65 in each of the 27 participating countries.The goal of PIAAC is to assess and compare the basic skills and competencies of adults around the world. The assessment focuses on cognitive and workplace skills needed for successful participation in 21st-century society and the global economy. Specifically, PIAAC measures relationships between individuals' educational background, workplace experiences and skills, occupational attainment, use of information and communications technology, and cognitive skills in the areas of literacy, numeracy, and problem solving.

To avoid repetition within the handbook, some of the statistical terms and procedures that are referred to in multiple chapters of the handbook are defined in **Appendix A. Glossary of Statistical Terms.**

Appendix B describes the various ways in which NCES publications and data files may be obtained. It also provides the reader with information on how to obtain a license for restricted-use data files.

Appendix C provides a list of the web-based and standalone tools for use with each of the NCES surveys.

Appendix D contains a list of the website addresses for each of the NCES surveys.

Appendix E contains an index.

Chapter 1: Early Childhood Longitudinal Study (ECLS)

1. OVERVIEW

The Early Childhood Longitudinal Study (ECLS) program is one of the active longitudinal surveys sponsored by NCES. The ECLS program includes three cohorts: a birth cohort and two kindergarten cohorts (the kindergarten class of 1998–99 and the kindergarten class of 2010–11). The birth cohort study (ECLS-B) followed a sample of children born in 2001 from birth through kindergarten; the first kindergarten study (ECLS-K) followed a sample of children who were in kindergarten in the 1998–99 school year through the eighth grade; and the second kindergarten study (ECLS-K:2011) will follow a sample of kindergartners in the 2010–11 school year through the fifth grade. The ECLS provides a comprehensive and reliable dataset with information about the ways in which children are prepared for school and how children develop in relation to their family, early childhood and school environments.

EARLY CHILDHOOD LONGITUDINAL SAMPLE SURVEY: BIRTH COHORT AND KINDERGARTEN COHORT

ECLS collects data from:

- Children
- Parents/guardians
- Child care providers and preschool teachers
- Teachers
- School administrators

Purpose

The ECLS provides national data on (1) children’s status at birth and at various points thereafter; (2) children’s transitions to nonparental care, early education programs, and school; and (3) children’s experiences and growth through the eighth grade. These data enable researchers to test hypotheses about associations and interactions of a wide range of family, school, community, and individual variables on children’s development, early learning, and performance in school.

Components

The ECLS has three cohort studies—two *kindergarten* cohort studies (ECLS-K and ECLS-K:2011) and the *birth* cohort study (ECLS-B)—and each of these has its own components.

The Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011). The ECLS-K:2011 will collect data from children, their families, classroom teachers, special education teachers, school administrators, and care providers on children’s cognitive, social, emotional, and physical development. Information also will be collected on children’s home environment, home educational activities, school environment, classroom environment, classroom curriculum, teacher background, and before- and after-school care.

The Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). The ECLS-K collected data from children, their families, classroom teachers, special education teachers, school administrators, and student records. The various components are described below.

Direct child assessments. The direct child assessments covered several cognitive domains (reading and mathematics in kindergarten through eighth grade; general knowledge, consisting of science and social studies questions, in kindergarten and first grade; and science in third, fifth, and eighth grades); a psychomotor assessment(fall kindergarten only), including fine and gross motor skills; and height and weight measurements. Beginning with the third-grade data collection, children

reported on their own perceptions of their abilities and achievement, as well as their interest in and enjoyment of reading, math, and other school subjects. An English language proficiency screener, the Oral Language Development Scale (OLDS), was administered to children if school records indicated that the child's home language was not English. The child had to demonstrate a certain level of English proficiency on the OLDS to be administered the ECLS-K cognitive assessment in English. If a child spoke Spanish at home and did not have the English skills required for the ECLS-K battery, the child was administered a Spanish version of the OLDS, and the mathematics and psychomotor assessments were administered in Spanish. . The assessment for each cognitive domain included a routing test (to determine a child's approximate skill level) and second-stage tests that were tailored to different skill levels. In the eighth-grade data collection, children completed a student questionnaire after completing the routing test. The student questionnaire covered many topics about the child's school experiences, school-sponsored and out-of-school activities, self-perceptions of social and academic competence and interests, weight and exercise, and diet.

Parent interviews. Parents/guardians were asked to provide key information about their children and their families, such as the demographics of household members (e.g., age, relation to child, race/ethnicity), family structure (household members and composition), parent/guardian involvement at the school and with children's schoolwork, home educational activities, children's child care experiences, child health, parental/guardian education and employment status, and their children's social skills and behaviors.

Classroom teacher questionnaire. In the kindergarten collections, all kindergarten teachers with ECLS-K-sampled children were asked to provide information on their educational backgrounds, teaching practices, teaching experiences, and the classroom settings in which they taught. They also were asked to complete a child-specific questionnaire that collected information on each sample child's social skills and approaches to learning, academic skills, and education placements. This procedure continued in later waves of the study. However, modifications were made beginning with the spring-fifth grade data collection, where the teachers who were most knowledgeable about the child's performance in each of the core academic subjects (i.e., reading/language arts, mathematics, and science) provided the data pertinent to each child's classroom environment and instruction for the academic subject about which they were most knowledgeable. Teachers also provided information about their professional background.

Special education teacher questionnaire. In each spring data collection, the primary special education teachers of and special education staff (e.g., speech pathologists, reading instructors, audiologists) who worked with sample children receiving special education services in school were asked to complete questionnaires about the children's experiences in special education, as well as their own professional background. Items in the special education teacher questionnaires addressed topics such as the child's disability, Individualized Education Program (IEP) goals, the amount and type of services sampled children received, and communication with parents and general education teachers about the child's special education program and progress.

School administrator questionnaire. School administrators were asked about school characteristics (e.g., school type, enrollment, and student body composition), school facilities and resources, community characteristics and school safety, school policies and practices, school-family-community connections, school programs programs for particular populations (e.g., limited English proficient students), staffing and teacher characteristics, school governance and climate, and their own characteristics.

Student records abstract. . In each round of data collection except eighth grade, school staff members were asked to complete a student records abstract form for each sampled child after the school year closed. These forms were used to obtain information about the child's attendance record, the presence of an IEP, the type of language or English proficiency screening that the school used, and (in the kindergarten year collection) whether the child participated in Head Start prior to kindergarten. A copy of each child's report card was also requested.

School facilities checklist. This checklist was used to collect information about the (1) availability and condition of the selected school's facilities, such as classrooms, gymnasiums, and toilets; (2) presence and adequacy of security measures; (3) presence of environmental factors that may affect the learning environment; and (4) overall learning climate of the school. An additional set of questions on portable classrooms was added to the spring first-grade data collection.

The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). The ECLS-B, which began in October 2001, was designed to study children's early learning and development from birth through the fall of

the kindergarten year. Over the course of the study, data were collected from multiple sources, including birth certificates, children, parents, nonparental care providers, teachers, and school administrators. These components are described below.

Birth certificates. These records provided information on the date of birth, child's sex, parents' education, parents' race and ethnicity (including Hispanic origin), mother's marital status, mother's pregnancy history, prenatal care, medical and other risk factors during this pregnancy and complications during labor and birth, and child's health characteristics at birth (such as congenital anomalies and abnormal conditions of the baby and the baby's Apgar score).

Parent/guardian interviews. A parent/guardian interview was conducted in the children's home at each data collection point to capture information about the children's early health and development, their experiences with family members and other significant people in their lives, the parents/guardians as caregivers, the home environment, and the neighborhood in which they lived. In most cases, the parent/guardian interviewed was the child's mother or female guardian.

Child assessments. Beginning at 9 months, children participated in activities designed to measure important developmental skills in the cognitive, socioemotional, and physical domains.

Cognitive domain. The cognitive assessments at the 9-month and 2-year data collections assessed general mental ability, including problem solving and language acquisition. The Bayley Short Form-Research Edition (BSF-R), designed specifically for the ECLS-B, was utilized in the 9-month month and 2-year data collections and consists of selected items from the Bayley Scales for Infant Development (BSID-II).

The cognitive assessments at the preschool, kindergarten 2006, and kindergarten 2007 data collections assessed early reading and mathematics and consisted of items from the ECLS-K as well as other studies and instruments. Color knowledge also was assessed in the preschool data collection.

Socioemotional domain. The Nursing Child Assessment Teaching Scale (NCATS) was used in the 9-month collection to assess child-parent interactions. An attachment rating, the Toddler Attachment Sort-45 (TAS-45), was used in the second wave of data collection. A videotaped parent-child interaction (Two Bags Task) was also used in the second and third waves of data collection.

Physical domain. In the 9-month data collection, children's height, weight, and middle upper arm circumference were assessed; additionally, a measure of head circumference was taken for children born with very low birth weight. These physical measures were taken again at all follow-up data collections. Additionally, children's fine motor skills and gross motor skills were assessed at all data collections (using the BSF-R motor scale in the 9-month and 2-year data collections and the ECLS-K Bruininks-Oseretsky Test of Motor Proficiency and Movement Assessment Battery for Children in the preschool, kindergarten 2006, and kindergarten 2007 data collections).

Nonparental care and education providers. Individuals and organizations that provided regular care for a child were interviewed with the permission of the child's parents. They were asked about their backgrounds, teaching practices and experience, the children in their care, and children's learning environments. This information was collected from the 2-year data collection on. In the kindergarten 2006 and 2007 collections, a wrap-around care provider interview was used for those children who were in kindergarten and had a before- or after-school care arrangement.

Teacher questionnaires and school data. Once the children entered kindergarten, teachers provided information on their classrooms and on children's cognitive and social development. Information for the school each child attended was obtained from NCES's school universe data files—the Common Core of Data (CCD) for public schools and the Private School Universe Survey (PSS) for private schools.

Father questionnaires. Fathers (both resident and nonresident fathers) completed a self-administered questionnaire, which asked questions about the particular role fathers play in their children's lives; the questionnaire provided information about children's well-being, the activities fathers engage in with their children, and key information about fathers as caregivers. Both resident and nonresident father questionnaires were included in the collections when the children were 9 months old and 2 years old. The resident father questionnaire was included in the preschool collection. No father questionnaires were included in the kindergarten collections.

Periodicity

The ECLS-K collected data in the fall and spring of kindergarten (1998–99), the fall of first grade (1999) (data were collected from a 30 percent subsample in this round), and in the spring of first grade (2000), third grade (2002), fifth grade (2004), and eighth grade (2007).

As currently planned, the ECLS-K:2011 will collect data in the fall and the spring of kindergarten (2010–11), the fall and the spring of first grade (2011-12), and the springs of second grade (2013), third grade (2014), fourth grade (2015), and fifth grade (2016).

The ECLS-B collected data when the children were about 9 months old (2001–02), about 2 years old (2003), about 4 years old (the preschool collection) (2005), and in the fall of kindergarten (2006 and 2007). Note that because of age requirements for school entry, children sampled in the ECLS-B entered kindergarten in two different years. All study children were included in the kindergarten 2006 collection, regardless of their enrollment status or grade in school. The kindergarten 2007 collection included just a portion of the total ECLS-B sample: children who were not yet in kindergarten in the 2006 collection, children who were in kindergarten in the 2006 collection and were repeating kindergarten in the 2007 collection, and twins of children in these groups. The ECLS-B study ended with the kindergarten 2007 wave of collection.

2. USES OF DATA

The ECLS-K provides information critical to informing policies that can respond sensitively and creatively to diverse learning environments. In addition, the ECLS-K enables researchers to study how a wide range of family, school, community, and individual variables are associated with early success in school and later development. The longitudinal nature of the study enables researchers to study children's reading achievement, growth in mathematics, and knowledge of the physical and social worlds in which they live. It also permits researchers to relate trajectories of growth and change to variations in children's school experiences in kindergarten and the early grades.

Like the kindergarten cohort study, the ECLS-B has two goals, descriptive and analytic. The study provides descriptive data on children's health status at birth; children's experiences in the home, nonparental care, and school; and children's development and growth through first grade. The data collected in the ECLS-B can be used to explore the relationships between children's developmental outcomes and their family, health care, nonparental care, school, and community.

The longitudinal nature of the study enables researchers to study children's physical, social, and emotional growth and to relate trajectories of growth and change to variations in children's experience.

3. KEY CONCEPTS

Number right scores. These scores are the counts of raw number of items a child answered correctly. These scores are useful for descriptive purposes only for assessments that are the same for all children. They are not comparable across grades. In the ECLS-K, some assessment items were not included as part of the set of proficiency scores (see details below) because they did not follow a hierarchical pattern. For these items, several item cluster scores were reported for the reading (kindergarten through fifth grade) and science assessments (third and fifth grades). These are simple counts of the number right on small subsets of items linked to particular skills. Because they are based on very few items, their reliability is relatively low.

Item Response Theory (IRT) scale scores. The ECLS direct cognitive assessments employ a two-stage design. As such, within any given domain, children receive a routing set of items (stage 1) and then based on their performance proceed to a certain difficulty level (stage 2). Because not all children receive all items, the assessment scores in the ECLS studies are modeled using Item Response Theory (IRT). Based on children's performance on the items they received, an ability estimate (theta) is derived for each domain. The theta is used to derive other scores, such as scale scores, T-scores, and proficiency probability scores. The IRT scale scores represent estimates of the number of items children would have answered correctly if they had received all of the scored questions in a given content domain. They are reported in both the ECLS-K and ECLS-B. They are useful in identifying cross-sectional differences among subgroups in overall achievement levels and provide a summary measure of achievement useful for correlations analysis with status variables. The IRT scale scores are also used as longitudinal measures of overall growth. Gain scores may be obtained by subtracting children's scale scores at two points in time.

Standardized scores (T-scores). These scores are also IRT based. They provide norm-referenced measurements of achievement; that is, estimates of achievement level relative to the population as a whole. A high mean T-score for a particular subgroup indicates that the group's performance is high in comparison to that of other groups. A change in mean T-scores over time reflects a change in the group's status with respect to that of other groups.

Proficiency probability scores. These scores are IRT-based and provide information on proficiency in clusters of items of similar difficulty along the overall

scale. The scores measure the probability of mastery of each level and can take on any value between 0 and 1. Because each proficiency probability score targets a particular set of skills, they are ideal for studying the details of achievement. They are useful as longitudinal measures of change because they show not only the extent of gains, but also where on the achievement (or development) scale the gains are taking place.

Race/ethnicity. In the ECLS, new Office of Management and Budget guidelines were followed under which a respondent could select one or more of five dichotomous race categories. In addition, a sixth dichotomous variable was created for those who simply indicated that they were multiracial without specifying the race. Each respondent additionally had to identify whether the child was Hispanic. Using the six dichotomous race variables and the Hispanic ethnicity variable, a race/ethnicity composite variable was created. The categories were White, non-Hispanic; Black or African-American, non-Hispanic; Hispanic, race specified; Hispanic, no race specified; Asian; Native Hawaiian or other Pacific Islander; American Indian or Alaska Native; and more than one race specified, non-Hispanic.

Socioeconomic status (SES). The SES variable reflects the SES of the household at the time of data collection. The components used to create the SES variable were father/male guardian's education, mother/female guardian's education, father/male guardian's occupation, mother/female guardian's occupation, and household income. In the ECLS-K, each parent's occupation was scored using the average of the 1989 General Social Survey (GSS) prestige scores for the 1980 census occupational category codes that correspond to the ECLS-K occupation code. In the ECLS-B, each parent's occupation was scored using the average of the 1989 GSS prestige scores for the 2000 census occupational category codes covered by the ECLS-B occupation.

4. SURVEY DESIGN

Target Population

Representative samples of kindergartners and babies are studied longitudinally for 6 or more years. Kindergarten children enrolled during the 1998–99 school year are the baseline for the ECLS-K cohort; babies born during 2001 are the baseline for the ECLS-B cohort.[1] Kindergarten children enrolled in the 2010–11 school year are the baseline for the ECLS-K:2011 cohort.

Sample Design

The sample design is discussed separately for the kindergarten and birth cohorts.

Kindergarten cohort (ECLS-K). The ECLS-K followed a nationally representative cohort of children from kindergarten through eighth grade.

Base-year (i.e., kindergarten) survey. A nationally representative sample children enrolled in kindergarten programs during the 1998–99 school year was sampled for participation in the study. These children were selected from both public and private schools, offering both full-day and part-day kindergarten programs. The sample was designed to support separate estimates of public and private school kindergartners; Black, Hispanic, White, and Asian/Pacific Islander children; and children grouped by SES.

The sample design for the ECLS-K was a dual-frame, multi-stage sample. First, 100 primary sampling units (PSUs) were selected from an initial frame of 1,400 PSUs, representing counties or groups of contiguous counties. The 24 PSUs with the largest measures of size (where the measure of size is the number of 5-year-olds, taking into account a factor for oversampling Asian/Pacific Islander 5-year-olds) were designated as certainty selections and were set aside. The remaining PSUs were partitioned into 38 strata of roughly equal measures of size. The frame of noncertainty PSUs was first sorted into eight superstrata by metropolitan statistical area (MSA) status and by census region resulting in four MSA superstrata and four non-MSA superstrata. Within the four MSA superstrata, the variables used for further stratification were race/ethnicity (high concentration of Asian/Pacific Islander, Black, or Hispanic), size of class, and 1988 per capita income. Within the four non-MSA superstrata, the stratification variables were race/ethnicity and per capita income. Two PSUs were selected from each noncertainty stratum using Durbin's method. This method selects two first-stage units per stratum without replacement, with probability proportional to size and a known probability of inclusion. The Durbin method was used because it allows variances to be estimated as if the units were selected with replacement.

School selection occurred within the sampled PSUs. Public schools were sampled from a public school

[1] The ECLS-B target population excludes children who were born to mothers younger than age 15 and children who died or were adopted prior to the 9-month home visit. Over time, the target population excludes children who died or moved abroad permanently.

frame (the 1995–96 CCD), and private schools were sampled from a private school frame (the 1995–96 PSS). The school frame was freshened in spring 1998 to include newly opened schools that were not included in the CCD and PSS (as well as schools that were included in the CCD and PSS but that did not offer kindergarten, according to these sources). A school sample supplement was selected from the freshened frame. In fall 1998, approximately 23 kindergarten children were selected, on average, from each of the sampled schools. Asian/Pacific Islander children and private schools were oversampled.

For the base year of the ECLS-K, 22,670 children were eligible (17,780 in public schools and 2,890 in private schools).

Fall first grade. The fall first grade collection was designed to enable researchers to measure the extent of summer learning loss and the factors associated with such loss and to better disentangle the relationships of school and home characteristics with children's learning. Data collection was limited to 26.7 percent of the base-year children in 30 percent of the originally sampled ECLS-K schools; that is, a total of 5,650 (4,450 public school and 1,200 private school) children. Data collection was attempted for every eligible child (i.e., a base-year respondent) still attending the school in which he or she had been sampled during kindergarten. To contain the cost of collecting data for a child who transferred from the school in which he or she was originally sampled, a random 50 percent of movers (i.e., children who changed schools) were flagged to be followed for the fall first-grade data collection.

Spring first grade. This data collection targeted all base-year respondents. In addition, the spring student sample was freshened to include current first-graders who had not been enrolled in kindergarten in 1998–99 and, therefore, had no chance of being included in the ECLS-K base-year kindergarten sample. While all students still enrolled in their base-year schools were recontacted, only a 50 percent subsample of base-year sampled students who had transferred from their kindergarten school was followed for data collection. For the spring first grade, 18,080 children were eligible (14,250 public school and 3,840 private school children). Student freshening brought 170 first-graders into the ECLS-K sample.

Spring third grade. The sample of children for the spring third-grade data collection consisted of all children who were base-year respondents and children who were brought into the sample in the spring of first grade through sample freshening. Sample freshening was not implemented in third grade. While all students still enrolled in their base-year schools were recontacted, slightly more than 50 percent of the base-year sampled students who had transferred from their kindergarten school were followed for data collection. This subsample of students was the same 50 percent subsample of base-year movers flagged for following in the spring of first grade, with the addition of movers whose home language was not English (followed at 100 percent). For the spring third grade, 16,670 children were eligible[2] (13,170 public schools and 3,500 private school children).

Spring fifth grade. In fifth grade, four groups of children were not followed, irrespective of other subsampling procedures that were implemented. These were (1) children who became ineligible in an earlier round (because they had died or moved out of the country), (2) children who were subsampled out in previous rounds because they had moved out of their original schools and were not followed, (3) children whose parents emphatically refused to cooperate in any of the data collection rounds since the spring of kindergarten, and (4) children eligible for the third-grade data collection for whom there were neither first-grade nor third-grade data.

Of the remaining children, those who moved from their original schools during fifth grade or earlier were subsampled for follow-up. In order to contain the cost of data collection, the rate of subsampling was lower in fifth grade than it had been in previous years. The subsampling rates maximize the amount of longitudinal data available for key analytic groups. Children whose home language was not English (English language learners or ELLs) continued to be sampled at higher rates (between 15 and 50 percent for base-year ELL respondents, and between 15 and 75 percent for ELL children freshened in first grade).

For the spring fifth grade, 12,030 children were eligible[3] (9,570 in public schools and 2,460 in private schools).

A new feature of the fifth-grade sample was the subsampling of eligible children for the administration of mathematics and science questionnaires. While all children retained for the fifth-grade data collection had child-level questionnaires filled out by their reading teachers, half had child-level questionnaires filled out by their mathematics teachers and the other half had

[2] This number reflects the longitudinal sample and excludes the 170 first grade freshened cases.
[3] This number reflects the longitudinal sample and excludes the 170 first grade freshened cases.

child-level questionnaires filled out by their science teachers.

Spring eighth grade. In the eighth-grade sample, the ineligible children were those who had moved out of the country, were deceased, or had moved to another school and were not subsampled for follow-up in an earlier grade. In the eighth-grade data collection, there was no subsampling of movers for follow-up as in previous rounds, since the majority of children did not remain in the same school from fifth grade to eighth grade (having moved out of elementary school into middle school).

For the spring eighth grade, 11,930 children were eligible[4] (9,480 in public schools and 2,450 in private schools).

Birth cohort (ECLS-B). The ECLS-B followed a nationally representative sample of children born in 2001 from the time the children were 9 months old through their kindergarten year.

Base-year (i.e., 9-month) survey. The ECLS-B sampled approximately 14,000 babies born in 2001, yielding approximately 10,700 completed cases in the 9-month collection. The sample included children from different racial/ethnic and socioeconomic backgrounds. Chinese children, other Asian/Pacific Islander children, children born with moderately low birth-weight (1,500–2,500 grams), children born with very low birth-weight (under 1,500 grams), and twins were oversampled. There was also a special supplemental component to oversample American Indian children.

The ECLS-B sample design consisted of a two-stage sample of PSUs and children born in the year 2001 within sampled PSUs. The PSUs were MSAs, counties, or groups of counties. Among the 96 sampled PSUs, 24 were large enough to be selected with certainty. The remaining PSUs were selected from groups of PSUs that were stratified by census region; MSA status; minority status (high/low); median income (high/low); and a composite measure of size, which was the expected number of births in 2001 in the PSU. Two PSUs were selected per stratum with probability proportional to size, a function of the expected number of births occurring within the PSU in 2001.

Births were sampled by place of occurrence, rather than by place of current residence. As a result, a different PSU sample than the PSU sample used in the ECLS-K, which uses residence-based population data, had to be selected. Within the sampled PSUs, children born in the year 2001 were selected by systematic sampling from birth certificates using the National Center for Health Statistics vital statistics record system. The sample was selected on a flow basis, beginning with January 2001 births (who were first assessed 9 months later, in October 2001). Approximately equal numbers of infants were sampled in each month of 2001. Different sampling rates were used for births in different subgroups, as defined by race/ethnicity, birth weight, and plurality (that is, whether or not the sampled newborn was a twin).

The sample of American Indian/Alaska Native (AI/AN) newborns drew from an additional 18 PSUs selected from a supplemental frame consisting of areas where the population has a higher proportion of AI/AN births. These PSUs were located in the western region of the United States. Six of the PSUs were selected with certainty. The noncertainty PSUs were selected independently of the core sample PSUs, with probability proportional to the number of AI/AN births.

Due to state-imposed operational restrictions and passive and active consent procedures, certain sampled PSUs had low expected response rates. For states where expected response rates were only slightly lower than planned, a larger sample was selected in order to achieve adequate numbers of respondents. Substitutions were made for PSUs in states where very low response rates were expected. The original PSU was matched with potential substitute PSUs on the criteria of median income; percentage of newborns in poverty; percentage of newborn Black, Hispanic, and other race/ethnicity children; population density; and birth rate. (AI/AN PSUs also were matched on tribal similarity. A Mahalanobis distance measure of similarity was used to create initial rankings.) Sampling rates from the original PSU were applied within the substitute PSU to obtain the original expected yield. A total of seven PSUs were used as substitutes for the original ECLS-B PSUs. Also, in two instances, an alternative frame was used to draw a sample of births occurring within PSUs with enrollment restrictions. Specifically, birth records were selected directly from hospital lists of births in counties that defined these original PSUs.

For the 9-month collection, approximately 14,200 children were eligible,[5] and approximately 10,700 participated.

[4] This number reflects the longitudinal sample and excludes the 170 first grade freshened cases.

[5] Because the ECLS-B data are restricted-use only, the numbers provided in this section for the ECLS-B are all rounded to the nearest 50.

Two-year collection. Only cases with a completed 9-month parent interview (about 10,700) were eligible for inclusion in the 2-year data collection. However, from that 10,700, about 100 cases where the child had died or moved abroad permanently between the 9-month and 2-year rounds were considered ineligible. There was no further sampling of cases. For the 2-year round of the ECLS-B approximately 9,850 cases participated (i.e. had a completed parent survey).

Preschool collection. All 9,850 cases with a complete 2-year parent interview and an additional 50 AI/AN cases were fielded and considered eligible for the preschool data collection, with the exception of approximately 100 cases in which children had died or moved permanently abroad between the 2-year interview and the preschool wave. For the preschool round of the ECLS-B approximately 8,950 cases participated (i.e., had a completed parent survey).

Kindergarten 2006 collection. For budgetary reasons, the kindergarten 2006 data collection followed a reduced sample (approximately 85 percent) of children who were eligible for the wave. The subsample was allocated disproportionately to the race/ethnicity, birth weight, and plurality domains to maintain larger sample sizes for the smaller domains. AI/AN children and Chinese children who were eligible were included with certainty in the kindergarten 2006 subsample. Eligible children were those with a parent response at all of the prior waves (9 months, 2 years, and preschool) and children sampled in the AI/AN domain with a parent response to the 9-month wave and at least one of the 2-year or preschool waves. AI/AN children who did not respond to either the 2-year or preschool waves were not included in the kindergarten 2006 wave. In addition, children who were identified as ineligible because they had died or moved out of the United States were not included in the kindergarten 2006 data collection.

After subsampling, approximately 7,700 children were eligible for the kindergarten 2006 wave and 7,000 participated (i.e., had a completed parent survey).

Kindergarten 2007 collection. The kindergarten 2007 data collection included a subset of the ECLS-B sample children with a completed parent interview at kindergarten 2006 and who met one of the following conditions: the child had not started kindergarten at the time of the kindergarten 2006 data collection; the child was the twin of a child who had not started kindergarten at the time of the kindergarten 2006 data collection; the child was in kindergarten during the kindergarten 2006 data collection and repeating kindergarten in school year 2007–08; or the child was the twin of a child who was repeating kindergarten in school year 2007–08.

Of the 7,000 cases from the kindergarten 2006 collection, based on the aforementioned criteria, 2,050 were eligible for the kindergarten 2007 (1,770 as first time entering school and 280 as likely repeating kindergarten). For the kindergarten 2007 wave, approximately 1,900 participated (i.e., had a completed parent survey).

Assessment Design

The design of the ECLS assessments is discussed separately for the kindergarten and birth cohorts.

Kindergarten cohort (ECLS-K). The design of the ECLS-K assessment was guided by the domain assessment framework proposed by the National Education Goals Panel's Resource Group on School Readiness. A critical component of the ECLS-K is the assessment of children along a number of dimensions, such as physical development, social and emotional development, and cognitive development. These domains were chosen because of their importance to success in school. The ECLS-K monitored the status and growth of its children along these domains:

- *Physical and psychomotor development:* Children's height and weight were measured at each data collection point in the ECLS-K. The psychomotor component was included only in the fall kindergarten collection. In that collection, kindergartners were asked to demonstrate their fine and gross motor skills through activities such as building a structure using blocks, copying shapes, drawing figures, balancing, hopping, skipping, and walking backward. Parents and teachers reported on other related issues, such as general health, nutrition, and physical activity. Beginning in third grade, the children also were asked to provide information about their eating habits and physical activity.

- *Social and emotional development:* The ECLS-K assessments of social and emotional development focused on the skills and behaviors that contribute to social competence. Aspects of social competence include social skills (e.g., cooperation, assertion, responsibility, self-control) and problem behaviors (e.g., impulsive reactions, verbal and physical aggression). Parents and teachers were the primary sources of information on children's social competence and skills in kindergarten and first grade. The measurement of children's social and emotional

development at grades three, five, and eight included instruments completed by the children themselves along with data reported by parents and teachers.

- *Cognitive development:* In kindergarten and first grade, the ECLS-K focused on three broad areas of competence: language and literacy, mathematics, and general knowledge of the social and physical worlds. Starting in third grade, a science assessment replaced the general knowledge assessment. In the higher grades, children's cognitive skills were expected to have advanced beyond the levels covered by the kindergarten and first-grade assessments; for this reason, a new set of assessment instruments was developed for third grade, for fifth grade, and again for eighth grade. Some of the assessment items were retained from one round to the next to support the development of longitudinal score scales in each subject area. The skills measured in each of these domains are a sample of the typical and important skills that are taught in American elementary schools and that children are expected to learn in school. The ECLS-K was developed to describe the behaviors, skills, and knowledge within broad cognitive domains that are most relevant to school curricula at each grade level and to measure children's growth from kindergarten to eighth grade. The ECLS-K assessment framework was based on current curricular domain frameworks for reading, mathematics, science, and social studies, as well as on existing assessment frameworks, such as those used in the National Assessment of Educational Progress. (See chapter 18.)

  The cognitive assessments were developed through extensive field testing and analysis of item performance. The final items were selected based on their psychometric properties and content relevance. Children's knowledge and skills in the natural and social sciences were measured in the general knowledge subdomain in kindergarten and first grade. The contents of this subtest, classified as science and social sciences, surveyed children's knowledge and understanding of relevant concepts. The science assessment used from third grade on measured children's knowledge in life science, physical science, and Earth science.

- Each direct cognitive domain subtest consisted of a routing test and second-stage tests that were tailored to different skill levels. All children were first administered a short routing test of domain-specific items having a broad range of complexity or difficulty levels. Performance on the routing test was used to determine the appropriate second-stage assessment form to be administered next to the child. The use of multilevel forms for each domain subtest minimized the chances of administering items that were all very easy or all very difficult for a given child. The assessments were administered in one-on-one, untimed sessions with a trained child assessor. If necessary, the session could take place over multiple periods.

Birth cohort (ECLS-B). The ECLS-B direct child assessment relied on instruments considered "gold standards" in the field. However, adaptations were necessary to take these instruments from a laboratory or clinic setting to a home setting. The ECLS-B child assessment was designed for ease of and flexibility in administration while at the same time being psychometrically and substantively sound. The key instruments used in the study were a shortened research edition of the BSID-II, called the Bayley Short Form-Research Edition (BSF-R), the NCATS, the Two Bags Task, an attachment measure—the TAS-45, and Bruininks-Oseretsky Test of Motor Proficiency and Movement Assessment Battery for Children.

- *Cognitive development and fine and gross motor skills:* The BSID-II is considered the gold standard for assessing early childhood development (ages 1 to 42 months). In the 9-month and 2-year collections, children's cognitive development, as well as their receptive and expressive language skills, were assessed using an adaptation of the mental scale of the BSID-II. Children retrieved hidden toys and looked at picture books, and their production of vowel-consonant combinations was noted. Fine and gross motor skills were assessed using an adaptation of the motor scale of the BSID-II. Children grasped small objects and were observed crawling and walking. The study had intended to field the entire Bayley assessment, as it was originally expected to take about 20 minutes to complete. However, a field test of the 9-month ECLS-B data collection revealed that it actually required an average of 40 minutes to complete. As a result, modifications were implemented to the original BSID-II. The ECLS-B contractor, Westat, worked with experts to identify a reduced-item set that could be administered in less time and could produce reliable, valid scores equivalent to the full set of Bayley items. The BSF-R took

approximately 25 minutes to administer. Because the BSF-R was not appropriate for children older than 42 months of age, a new direct child cognitive assessment was developed for use in the preschool and kindergarten collections. These assessments were patterned after the ECLS-K assessments and incorporated items from the ECLS-K, as well as other published assessments, such as the preLas 2000, Test of Early Mathematics Ability, Third Edition (TEMA 3), and the Peabody Picture Vocabulary Test, Third Edition (PPVT-III). The cognitive domains covered in the preschool-kindergarten assessments were early reading and mathematics skills. The preschool collection also included a measure of children's color knowledge, which involved asking the children to name the colors of each bear presented to them in picture format. Children's fine and gross motor skills were measured using the Bruininks-Oseretsky Test of Motor Proficiency and Movement Assessment Battery for Children. To assess fine motor skills, children were asked to copy a series of forms (e.g., circle, triangle, square) that were first drawn by an assessor and to build a structure with blocks that was first demonstrated by the assessor. To assess gross motor skills, children were asked to hop, skip, jump backwards, and balance on one foot.

Because the NCATS is only appropriate for children up to 36 months of age, the Two Bags Task was used in the 2-year and preschool data collections. The Two Bags Task is a simplified version of the Three Bags Task that was used successfully in such large-scale studies as the Early Head Start Research and Evaluation Project and is intended to capture children's socioemotional functioning. It is a semistructured activity completed by the parent and child in interaction. During this 10-minute task, the parent-child dyad is asked to play with two different sets of toys, each placed within a separate numbered bag. In the 2-year collection, bag number 1 contained a children's picture book and bag number 2 contained a set of dishes. In the preschool collection, bag number 1 also contained a children's picture book but bag number 2 contained PlayDoh. The rating scales provide information on parents' behaviors during the interaction (parental sensitivity, intrusiveness, simulation of cognitive development, positive regard, negative regard, and detachment) and children's behaviors during the interaction (child engagement of parent, sustained attention, and negativity toward parent).

In the preschool and kindergarten collections, information on children's socioemotional functioning was collected indirectly through questions asked of parents and teachers.

- *Children's security of attachment:* The TAS-45 is a modified version of the Attachment Q-Sort (AQS), a widely used observational measure of children's security of attachment. It includes 45 items describing children's behaviors. After being in the home with the child and parent for several hours, the ECLS-B assessors completed a task in which they indicated whether each of the 45 behaviors applied to the child and how strongly the behavior either applied or did not apply, based upon their observations of the child in the home. These items/behaviors cluster around common attachment-related constructs, such as "cooperativeness," "independence," or "attention-seeking." Nine clusters, or "hot spots," were identified in the data. These hot spots, along with a traditional attachment classification (Avoidant, Secure, Ambivalent, and Disorganized) and traditional security and dependency scores were developed from the TAS-45. The TAS-45 was only administered in the 2-year data collection.

Data Collection and Processing

The ECLS-K compiled data from four primary sources: children, children's parents/guardians, teachers, and school administrators. Data collection began in fall 1998 and continued through spring 2007. Self-administered questionnaires, one-on-one assessments, and telephone or in-person interviews were used to collect the data. Westat conducted all rounds of data collection from kindergarten through eighth grade.

The ECLS-B compiled data from multiple sources, including administrative records, children, parents, nonparental care providers, teachers, and NCES school universe files. Data collection began in 2001 and continued through 2008. The primary modes of data collection were an in-person home visit during which parent respondents were interviewed and children were directly assessed. Self-administered questionnaires and telephone interviews also were used to collect data. Westat was the 9-month and 2-year data collection contractor. RTI International conducted the preschool and kindergarten data collections.

Reference dates. For the ECLS-K, baseline data for the fall were collected from September through December

1998. For the ECLS-B, baseline data were collected from October 2001 through December 2002.

Data collection. The ECLS-K and the ECLS-B are discussed separately.

Kindergarten cohort (ECLS-K). The data collection schedule for the ECLS-K was based on a desire to capture information about children as critical events and transitions were occurring rather than measuring these events retrospectively. A large-scale field test of the kindergarten and first-grade assessment instruments and questionnaires was conducted in 1995–96. This field test was used primarily to collect psychometric data on the ECLS-K assessment item pool and to evaluate questions in the different survey instruments. Data from this field test were used to develop the routing and second-stage tests for the ECLS-K kindergarten and first-grade direct cognitive assessment battery and to finalize the parent, teacher, and school administrator instruments. A pilot test of the systems and procedures, including field supervisor and assessor training, was conducted in April and May 1998 with 12 elementary schools in the Washington, DC, metropolitan area. Modifications to the data collection procedures, training programs, and systems were made to improve efficiency and reduce respondent burden. Modifications to the parent interview to address some issues raised by pilot test respondents were also made at this time.

Data on the kindergarten cohort were collected twice during the base year of the study—once in the beginning (fall) and once near the end (spring) of the 1998–99 school year. The fall 1998 data collection obtained baseline data on children at the very beginning of their exposure to the influences of school, providing measures of the characteristics and attributes of children as they entered formal school for the first time. The data collected in spring 1999, together with the data from the beginning of the school year, are used to examine children's first encounter with school. Data were collected from the child, the child's parents/ guardians, and teachers in both fall and spring. Data were collected from school administrators in the spring. For the fall 1998 and spring 1999 collections, all child assessment measures were obtained through untimed CAPI, administered one-on-one by the assessor to the child. The assessment was normally conducted in a school classroom or library and took approximately 50 to 70 minutes per child. Children with a primary home language other than English (according to school records) were first administered an English language screener (OLDS) to determine whether their English language skills were sufficient enough to take the cognitive assessments in English. Children who fell below the cut score for the OLDS and whose language was Spanish were administered a Spanish-language version of the OLDS and the ECLS-K mathematics assessment translated into Spanish, and they had their height and weight measured. Children who fell below the cut score and whose language was neither English nor Spanish had only their height and weight measured. (A child was administered the OLDS in each round of data collection until he or she passed it; the OLDS was no longer used after the spring first grade data collection because by then most children demonstrated sufficient English language skills to be assessed in English.) Most of the parent data were collected by computer-assisted telephone interviewing (CATI), though some of the interviews were collected through CAPI when respondents did not have a telephone or were reluctant to be interviewed by telephone. All kindergarten teachers with sampled children were asked to fill out self-administered questionnaires providing information on themselves and their teaching practices. For each of the sampled children they taught, the teachers also completed a child-specific questionnaire. In the spring, school administrators were asked to complete a self-administered questionnaire that included questions on the school characteristics and environment, as well the administrator's own background. Also, in the spring, the special education teachers or service providers of children in special education were asked to complete a self-administered questionnaire about the children's experiences in special education and about their own background. In addition, school staff members were asked to complete a student record abstract after the school year closed.

In fall 1999, when most of the kindergarten cohort had moved on to first grade, data were collected from a 30 percent subsample of the cohort. The direct child assessment was administered during a 12-week field period (September–November 1999). The parent interview was administered between early September and mid-November 1999; it averaged 35 minutes, and was conducted primarily by telephone.

Spring data collections (first grade, third grade, fifth grade, and eighth grade) included direct child assessments, parent interviews, teacher and school questionnaires, student record abstracts, and facilities checklists. As in other rounds, the child assessments were administered with CAPI (March–June 2000 for the first-grade collection, March–June 2002 for the third-grade collection, February–June 2004 for the fifth-grade collection, and March–June 2007 for the eighth-grade collection), while both CATI and CAPI were used for the parent interview (March–July 2000 for first grade, March–July 2002 for third grade, February–June 2004 for fifth grade, and March–June

2007 for eighth grade). Self-administered questionnaires were used to gather information from teachers, school administrators, and student records (March–June 2000 for first grade and March–June 2002 for third grade, but field staff prompted by telephone for the return of these materials through October 2000 and October 2002, respectively. For the fifth grade, data collection was between February and June 2004. For the eighth grade, data collection was between March and June 2007.).

A continuous quality assurance process was applied to all data collection activities. Data collection quality control efforts began with the development and testing of the CATI and CAPI applications and the contractor's Field Management System. As these applications were programmed, extensive testing of the system was conducted. Quality control processes continued with the development of field procedures that maximized cooperation and thereby reduced the potential for nonresponse bias. Quality control activities also were practiced during training and data collection. During the original assessor training, field staff practiced conducting the parent interview in pairs and practiced the direct child assessment with kindergarten children brought to the training site for this purpose. In later data collection periods, experienced staff used a home study training package while new staff were trained in classroom sessions. After data collection began, field supervisors observed each assessor conducting child assessments and made telephone calls to parents to validate the interview. Field managers also made telephone calls to the schools to collect information on the school activities for validation purposes.

Birth cohort (ECLS-B). A field test of the ECLS-B instruments and procedures was conducted in the fall of 1999. The design featured many different tasks. For example, while in the home, a field staff member had to complete approximately 11 discrete tasks, and each task had special skill requirements. Early in the field test, NCES and the ECLS-B contractor found several problems regarding the complexity of the home visit: while separately no one task was difficult, the total data collection protocol was complex, so it was necessary to simplify these tasks in order to reduce the burden on field staff and to ensure the reliable and valid administration of all tasks. As a result, several modifications were made to the original data collection design.

A second field test of the ECLS-B instruments and procedures began in September 2000. A field test sample was drawn consisting of 1,060 children born between January and April 2000. Home visits were conducted when the children were 9 months old and again when they were 18 months old. Results from this field test indicated that the changes to the design that resulted from the first field test were successful.

The ECLS-B schedule called for information to be gathered on the children and from the parents during an in-home visit. The children's mother or primary caregiver was the respondent for the parent interview at each round of data collection. Child assessments were conducted in the child's home by the trained ECLS-B assessors at every round of data collection as well. Resident fathers (defined as the spouse or partner of the female parent respondent) were asked to complete a self-administered questionnaire with questions regarding their involvement in their children's lives in the 9-month, 2-year, and preschool data collections. Biological, non-resident fathers were asked to complete a self-administered questionnaire in the 9-month and 2-year data collections if the mother gave permission for him to be contacted. In the 2-year and preschool data collections, information was collected from children's primary nonparental care providers through a telephone interview. Direct observations to assess child care quality also were conducted by trained observers for a subsample of children with regular nonparental care. In the kindergarten 2006 collection, the child care provider telephone interview used in the preschool collection was again fielded for children who had not yet entered kindergarten. A wrap-around care and education provider telephone interview (WECEP) was introduced in this collection to obtain information on children's before- and after-school care arrangements for those children who were in kindergarten. The WECEP was used in the kindergarten 2007 collection as well. Observations of care settings were not conducted in the kindergarten collections. Teachers of children in kindergarten in 2006 and 2007 were asked to complete a self-administered questionnaire similar to those used in the ECLS-K that asked about the child's classroom, the child's behaviors and performance in the classroom, and their own background. Although the ECLS-B did not include a school administrator questionnaire, information on children's schools was obtained from the NCES school universe files, the Common Core of Data (CCD) for public schools and the Private School Survey (PSS) for private schools.

The ECLS-B 9-month data collection began in October 2001 and continued through December 2002. The 2-year data collection began in January 2003 and continued through April 2004. While the 9-month and 2-year data collection schedules were designed to collect information on children as close as possible to the date on which they turned the age of interest for the collection (i.e., 9 months and 2 years), the collection

schedules for the preschool and kindergarten rounds were changed to correspond with an academic calendar. Thus, the preschool wave of data collection began in late August 2005 and ended in mid-July 2006. The kindergarten 2006 collection began in fall 2006 through spring 2007. The kindergarten 2007 collection began in fall 2007 through spring 2008. In all collections, CAPI was the principal mode of data collection for the parent interview, Self-administered questionnaires were used to gather information from the resident father, nonresident father, and teacher. A self-administered questionnaire was used to obtain information on potentially sensitive topics from the parent respondent at 9 months and 2 years; starting with the preschool collection, potentially sensitive items were administered using audio computer-assisted self-interviewing technology (ACASI). Data were collected from the child by several means: a series of structured, standardized activities were scored in the home by the field interviewer; structured interactions with the parent were videotaped for later coding; physical measurements were obtained; and behavior was observed throughout the home visit.

Child-parent interactions were assessed by NCATS at the 9-month data collection, and again by the Two Bags Task at the 2-year and preschool data collections. In all cases, the ECLS-B videotaped these structured interactions. Although it is more typical for a health or social service professional to complete NCATS via live coding (i.e., while the interaction is occurring), the ECLS-B field staff needed to observe and score 73 items of parent and child behavior. Given the other tasks the field staff had to learn and complete, live coding would have limited the number of scales that could realistically be used, thereby reducing the amount of information that could be gathered. The videotapes were coded along all scales.

Data were collected from child care providers by means of CATI. A subset of child care providers was sampled for on-site observations in the 2-year and preschool collections; observers recorded data in booklets, and child care center directors completed a self-administered paper questionnaire.

Editing. Within the CATI/CAPI instruments, the ECLS-K and ECLS-B respondent answers were subjected to both “hard” and “soft” range edits during the interviewing process. Responses outside the soft range of reasonably expected values were confirmed with the respondent and entered a second time. For hard-range items, out-of-range values were usually not accepted. If the respondent insisted that a response outside the hard range was correct, the assessor could enter the information in a comments data file. Data preparation and project staff reviewed these comments. Out-of-range values were accepted if the comments supported the response.

Consistency checks were also built into the CATI/CAPI data collection. When a logical error occurred during an interview, the assessor saw a message requesting verification of the last response and a resolution of the discrepancy. In some instances, if the verified response still resulted in a logical error, the assessor recorded the problem either in a comment or in a problem report.

The overall data editing process consisted of running range edits for soft and hard ranges, running consistency edits, and reviewing frequencies of the results. Where applicable, these steps also were implemented for hard-copy questionnaire instruments, videotaped instruments, and observational instruments.

Estimation Methods

Data were weighted to account for differential probabilities of selection at each sampling stage and to adjust for the effects of nonresponse. A hot-deck imputation methodology was used to impute missing values for all components of SES in the ECLS-K and ECLS-B. Imputation also was implemented for child assessment proficiency-level variables and free/reduced-price school lunch data in the ECLS-K.

Weighting. Weighting in the ECLS-K and ECLS-B is discussed separately.

Kindergarten cohort (ECLS-K). Several sets of weights were computed for each of the seven rounds of data collection (fall kindergarten, spring kindergarten, fall first grade, spring first grade, spring third grade, spring fifth grade, and spring eighth grade). These weights include cross-sectional weights for analyses of data from one time point, as well as longitudinal weights for analyses of data from multiple rounds of the study. Unlike surveys that have only one type of survey instrument aimed at one type of sampling unit, the ECLS-K is a complex study with multiple types of sampling units, each having its own survey instrument. Each type of unit was selected into the sample through a different mechanism: children were sampled directly through a sample of schools; parents of the sampled children were automatically included in the survey; all kindergarten teachers and administrators in the sampled schools were included; and special education teachers were included in the sample if they taught any of the sampled children. Each sampled unit had its own survey instrument: children were assessed directly using a series of cognitive and physical assessments; parents were interviewed with a parent instrument;

teachers filled out at least two different types of questionnaires, depending on the round of data collection and on whether they were regular or special education teachers; and school principals reported their school characteristics using the school administrator questionnaire. The stages of sampling, in conjunction with different nonresponse levels at each stage and the diversity of survey instruments, required that multiple sampling weights be computed for use in analyzing the ECLS-K data.

Weight development was driven by three factors: (1) how many points in time would be used in analysis (i.e., whether the analysis would be longitudinal or cross-sectional); (2) what level of analysis would be conducted (e.g., child, teacher, or school); and (3) what source of data would be used (e.g., child assessment, teacher questionnaire, parent questionnaire).

For the kindergarten rounds of data collection, weights were computed in two stages. In the first stage, base weights were computed. The base weights are the inverse of the probability of selecting the unit. In the second stage, base weights were adjusted for nonresponse. Nonresponse adjustment cells were generated using variables with known values for both respondents and nonrespondents. Chi-squared Automatic Interaction Detector (CHAID) analyses were conducted to identify the variables most highly related to nonresponse. Once the nonresponse cells were determined, the nonresponse adjustment factors were calculated as the reciprocals of the response rates within the selected nonresponse cells. Beginning with the first grade round of data collection, a third stage called raking was introduced into the weight development process to remove the variability due to the subsampling of schools and children who changed schools (i.e., movers). In this stage, child weights were raked to sample-based control totals computed using the base year child weights adjusted for nonresponse.

The base weight for each school is the inverse of the probability of selecting the PSU in which the school is located multiplied by the inverse of the probability of selecting the school within the PSU. The base weights for eligible schools were adjusted for nonresponse; this was done separately for public and private schools.

The base weight for each child in the sample is the school nonresponse-adjusted weight for the school attended multiplied by a poststratified within-school student weight (total number of students in the school divided by the number of students sampled in the school). The poststratified within-school weight was calculated separately for Asian/Pacific Islander and non-Asian/Pacific Islander children because different sampling rates were used for these two groups. Within a school, all Asian/Pacific Islander children have the same base weights and all non-Asian/Pacific Islander children have the same base weights. The parent weight, for use with analysis of parent data, is the base child weight adjusted for nonresponse to the parent interview. Again, these adjustments were made separately for students in public and private schools. The teacher weight, for use with child-level analysis that includes teacher data from the child-level questionnaire specific to the sample child, is the base child weight adjusted for nonresponse to the teacher child-level questionnaire. Weights for child-level analysis were developed for every round of data collection. Weights for analysis at the school and teacher levels (i.e., weights that allow for the generation of national estimates of schools educating kindergarten-age children and kindergarten teachers) were developed only for the kindergarten data collections. The sample is not representative of schools or teachers after the kindergarten year,

Birth cohort (ECLS-B). Several sets of weights were computed for each round of data collection. Weights are used to adjust for disproportionate sampling, survey nonresponse, and noncoverage of the target population when analyzing complex survey data. The weights are designed to eliminate or reduce bias that would otherwise occur with analyses of unweighted data. The ECLS-B weights were developed in three steps: First, base weights were calculated using the overall selection probabilities; next, weights were adjusted for survey nonresponse; finally, raking was used to adjust for undercoverage and to improve the precision of survey estimates.

The base weight gives the approximate representation of each sampled birth record. The base weight for a given birth record was calculated as the reciprocal of the overall probability of selection, computed as the product of each stage's probability of selection. These overall probabilities of selection and base weights are used to compute analysis weights for all ECLS-B children in each round of data collection.

Next, base weights were adjusted for survey nonresponse. A selected set of variables related to child and family characteristics was used to construct nonresponse adjustment cells for each set of weights. Respondents and nonrespondents were compared on the characteristics selected based on analyses using segmentation modeling via CHAID. In the first round of data collection, data from the birth certificate were used to compare respondents and nonrespondents, because these data were available for all sampled cases regardless of participation status. In later collections,

respondents and nonrespondents were compared on both birth certificate data and data collected in prior rounds. A nonresponse adjustment factor was calculated for each cell as the ratio of the sum of weights for eligible cases in the cell to the sum of weights for eligible and responding cases in the cell. Finally, the nonresponse-adjusted weights were raked to 11 dimensions to ensure that sums of weights matched known population totals, thus correcting for survey undercoverage. The 11 dimensions were selected because of their substantive interest as well as their relationship to response propensity, as indicated by the CHAID modeling and also some preliminary logistic regression analyses.

The development of the ECLS-B weights was a sequential process. The 9-month weights were developed first, starting with the base weights; the 2-year weights were developed as adjustments to the 9-month weights; the preschool weights started with the 2-year weights, the kindergarten 2006 weights started with the preschool weights, and the kindergarten 2007 weights started with the kindergarten 2006 weights. A set of weights also was developed to allow for analysis of children in their first year of kindergarten, whether that year was in the 2006 collection or the 2007 collection. These weights were developed as adjustments to the preschool weights. As there are three main components in the 9-month round (parent interview data, child assessment data, and father data) and five or more components in each of the following rounds (parent interview data, child assessment data, father data, child care provider data, child care observation data, teacher data, and/or school data, depending on the round), several sets of weights were developed, taking into account the level of nonresponse for the different components and combinations of completed components that would be of most analytic interest. For example, the 9-month parent-father-child weight is valid for cases for which all three components are complete and adjusts for nonresponse to these components, whereas the 9-month parent weight is valid for all cases for which the parent component is complete, regardless of whether the child or father components are complete, and adjusts for nonresponse to the parent interview. Both cross-sectional weights for analysis of data at one round and longitudinal weights for analysis of data from multiple rounds of the study were computed.

Scaling. IRT was employed in the ECLS-K and ECLS-B to calculate scores that could be compared both within a round and across rounds, regardless of which second-stage form a student took. The items in the routing test, plus a core set of items shared among the different second-stage forms, made it possible to establish a common scale.

Imputation.
Kindergarten cohort (ECLS-K). In the ECLS-K, SES component variables were computed for the base-year, spring first-grade, spring third-grade, spring fifth-grade, and spring eighth-grade rounds. The percentages of missing data for the education and occupation variables were small (2 to 11 percent in the base year, 4 to 8 percent in the spring of first grade, 2 to 3 percent in the spring of third grade, 1 to 2 percent in the spring of fifth grade; and 3 percent in the spring of eighth grade); however, the household income variable had a higher rate of missing data (28.2 percent in the base year and 11 to 33 percent in the spring of first grade, depending on whether a detailed income range or the exact household income was requested; in the spring of third grade, 11.1 percent of cases had missing data for the detailed income range; this percentage was 8.1 percent of cases in the spring of fifth grade and 7.0 percent of cases in the spring of eighth grade). A standard (random selection within class) hot-deck imputation methodology was used to impute for missing values of all SES components in all years. From the spring of first grade on, the initial step in the imputation procedure was to fill in missing values from information gathered during an earlier interview with a parent, if one had taken place. If no prior data were available, standard hot-deck imputation was used.

The SES component variables were highly correlated, so a multivariate analysis was more appropriate to examine the relationship between the characteristics of donors and nonrespondents. For the base year, CHAID was used to divide the data into cells based on the distribution of the variable to be imputed, as well as to analyze the data and determine the best predictors. These relationships were used for imputation in later rounds of the ECLS-K.

The variables were imputed in sequential order and separately by type of household. For households with both parents present, the mother's and father's variables were imputed separately. If this was not the case, an "unknown" or missing category was created as an additional level for the CHAID analysis. As a rule, no imputed value was used as a donor. In addition, the same donor was not used more than two times. The order of the imputation for all the variables was from the lowest percentage missing to the highest.

Imputation for occupation involved two steps. First, the labor force status of the parent was imputed, whether the parent was employed or not. Then the parent's occupation was imputed only for those parents whose

status was identified as employed, either through the parent interview or the first imputation step. The variable for income was imputed last using a three-stage procedure; if a respondent provided partial information about income, this was used in the imputation process.

Imputation was also employed for variables related to the percentage of children in a school who received free or reduced-price lunch. Not all school principals answered all three questions that were used to derive the composite variables indicating the percentage of students in the school who received free lunch and the percentage who received reduced-price lunch: total school enrollment, number of children eligible for free lunch, and number of children eligible for reduced-price lunch. Prior to the fifth grade, if these three source variables had missing values, the composites were filled in with values computed using the most recent CCD data if they were not missing from the CCD, or left missing if they were missing from the CCD. Beginning in fifth grade, missing values in the composite variables were imputed. Missing values in the source variables, however, were not imputed.

A two-stage procedure was used for imputing school lunch composites. First, if a school had nonmissing values for the school lunch composites in kindergarten, first grade, and third grade, missing values for the fifth grade were filled in with values from previous years. A similar procedure was employed for eighth grade, which was first if a school had nonmissing values for a prior round, eighth grade was filled with the value from the previous year. Second, data still missing after this initial step were imputed using a hot-deck methodology. Imputation cells were created using the Title I status of the school and school longitude and latitude. School data that were imputed by hot deck are generally transfer schools with few sample children.

Birth cohort (ECLS-B). As in the ECLS-K, variables used to derive the SES composite variable were imputed using a hot-deck methodology. These variables include mother’s and father’s education, mother’s and father’s occupation, and income range. Imputation cells were defined by respondent characteristics that were the best predictors of the variables to be imputed, as determined using a CHAID analysis. Hot-deck imputation was done in a sequential order, separately, by type of household (female single parent, male single parent, and both parents present). As with the ECLS-K, missing data from a later round were first filled with data obtained in a prior round, if available. For households with both parents present, the mother’s and father’s variables were imputed separately. Imputed as well as reported values were used to define imputation cells; missing values for donor characteristics were treated as a separate category. No imputed value was used as a donor. No donor was used more than once. The order of hot-deck imputation for all variables was from the lowest percentage missing to the highest.

Future Plans

The ECLS-K:2011 will follow students from kindergarten in 2010 through fifth grade in 2015. Because it is designed to allow for comparisons between the 2010-11 cohort and the cohort of kindergartners included in the ECLS-K, by design the ECLS-K:2011 is very similar to the ECLS-K and includes most of the same components. Some changes of note are the introduction of a basic reading skills assessment to be administered to all children, regardless of primary home language; a Spanish basic reading skills assessment to be administered to Spanish-speaking children who do not pass an English language screener; and the replacement of the final and gross motor skills assessments with an assessment of children’s executive functions, a set of interdependent processes that work together to accomplish purposeful, goal-directed activities and include working memory, attention, inhibitory control, and other self-regulatory processes.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

The estimators of sampling variances for the ECLS statistics take the ECLS complex sample design into account. Both replication and Taylor Series methods can be used to accurately analyze data from the studies. The paired jackknife replication method using 90 replicate weights can be used to compute approximately unbiased estimates of the standard errors of the estimates. (The fall first-grade subsample in the ECLS-K uses 40 replicate weights.) When using the Taylor Series method, a different set of stratum and first-stage unit (i.e., PSU) identifiers should be used for each set of weights. Both replicate weights and Taylor series and identifiers are provided as part of the ECLS-K and ECLS-B data files.

Design effects.

Kindergarten cohort (ECLS-K). A large number of data items were collected from students, parents, teachers, and schools. Each item has its own design effect that can be estimated from the survey data. The median child-level design effect is 4.7 for fall kindergarten and 4.1 for spring kindergarten. The median child-level

design effect for spring third grade, spring fifth grade, and spring eighth grade is 3.3, 4.0, and 3.1, respectively.

The size of the ECLS-K design effects is largely a function of the number of children sampled per school. With about 20 children sampled per school, an intraclass correlation of 0.2 might result in a design effect of about 5. The median design effect is 3.4 for the panel of students common to both the fall and spring of kindergarten, and the lower median design effect is due to the smaller cluster size in the panel. The ECLS-K design effects are slightly higher than the average of 3.8 (with the exception of the spring third-grade collection and spring eighth-grade collection design effect) that was anticipated during the design phase of the study, both for estimates for proportions and for score estimates.

The median teacher-level design effect is 2.5 for both the fall and spring of kindergarten. This design effect is lower than the child-level design effects because the number of responding teachers per school is relatively small. The design effect for teachers is largely a result of selecting a sample using the most effective design for child-level statistics, rather than a design that would be most effective for producing teacher-level statistics.

The median school-level design effect is 1.6. Design effects were not computed for items from the teacher and school administrator questionnaires in the spring of first, third, fifth, and eighth grades because no teacher or school weights were computed for any of the ECLS-K years after kindergarten.

A multilevel analysis was carried out to estimate components of variance in the fall- and spring-kindergarten cognitive scores associated with (1) the student, (2) the school, (3) the data collection team leader, and (4) the individual test administrator. This secondary analysis was motivated by Westat's earlier finding of larger-than-expected design effects. In addition, the impact of parent's education on the above sources of variance was also estimated.

Birth cohort (ECLS-B) As noted above, several sets of weights were developed for use with different combinations of survey components that are of analytic interest. Design effects were computed for different survey estimates produced using these different weights. Using the parent weights, the median parent-level design effect is 2.1 for the 9-month data collection, 2.4 for the 2-year collection, 2.1 for the preschool collection, 2.0 for the kindergarten 2006 collection, and 2.2 for the kindergarten 2007 collection. The median design effects for other weights across all components and all rounds of collections ranges from a low of 1.2 for the 2-year weight connected to response to the child care observation (W22P0) weight and a high of 4.2 for the 9-month weight connected to response to the 9-month child assessment (W1C0) weight.

It is noted that the design effects for assessment estimates are higher than the design effects for some other types of estimates. This can be due to either naturally occurring higher intracluster correlations for assessment estimate items or interviewer effects. In the ECLS-B, where the general relationship between interviewer and cluster is one-to-one, the two are difficult, if not impossible, to disentangle. Similar observations about the design effects for assessment estimates were made in the ECLS-K data.

Nonsampling Error

In order to reduce nonsampling error, the survey design phase included focus groups and cognitive laboratory interviews for the purposes of assessing respondent knowledge topics, comprehension of questions and terms, and item sensitivity. The design phase also entailed testing of the CAPI instrument and a field test that evaluated the implementation of the survey.

Another potential source of nonsampling error is respondent bias that occurs when respondents systematically misreport (intentionally or unintentionally) information in a study. One potential source of respondent bias in the ECLS surveys is social desirability bias. If there are no systematic differences among specific groups under study in their tendency to give socially desirable responses, then comparisons of the different groups will accurately reflect differences among the groups. An associated error occurs when respondents give unduly positive assessments about those close to them. For example, parents may give more positive assessments of their children's experiences than might be obtained from institutional records or from the teachers.

Potentially, response bias may also be introduced in the responses of teachers about each individual student. For example, each teacher filled out a survey for each of the sampled children they taught in which they answered questions on the child's socioemotional development in the ECLS-K and ECLS-B. Since the base-year and first-grade surveys in the ECLS-K and the kindergarten surveys in the ECLS-B were first conducted in the fall, it is possible that the teachers did not have adequate time to observe the children, and thus some of their responses may be influenced by their expectations based on which groups (e.g., sex, race,

ELL status, disability) the children belonged to. In order to minimize bias, all items were subjected to multiple cognitive interviews and field tests, and actual teachers were involved in the design of the cognitive assessment battery and questionnaires. NCES also followed the criteria recommended in a working paper on the accuracy of teachers' judgments of students' academic performances (see Perry and Meisels 1996).

As in any survey, respondent bias may be present in the ECLS-K and ECLS-B. It is not possible to state precisely how such bias may affect the results. NCES has tried to minimize some of these biases by conducting one-on-one, untimed assessments, and by asking some of the same questions about the sampled child of both teachers and parents.

Coverage error. Undercoverage occurs when the sampling frame used does not fully reflect the target population of inference. By designing the ECLS-K child assessment to be both individually administered and untimed, both coverage error and bias were reduced. Individual administration decreases problems associated with group administration, such as children slowing down and not staying with the group or simply getting distracted. The advantage of having untimed exams was that the study was able to include most children with special needs and/or who needed some type of accommodation, such as children with a learning disability, with hearing aids, etc. The only children who were excluded from the study were those who were blind, those who were deaf, those whose IEP clearly stated that they were not to be tested, and non-English-speaking children who were determined to lack adequate English or Spanish language skills to meaningfully participate in the ECLS-K battery. Exclusion from the direct child assessment did not exclude children from other parts of the study (e.g., teacher questionnaire, parent interview).

For the ECLS-B, the 9-month target population is all infants born in the United States in 2001 to mothers 15 years of age and older who were not adopted prior to, and who were alive during, the 9-month data collection period. The target population for later rounds of collection also excludes children who died or moved abroad permanently. Concern about noncoverage in the ECLS-B relates mainly to a few PSUs where births were sampled from hospital frames. In addition, the main sampling frame consisted of birth certificates available from state registrars. This sampling frame failed to cover unregistered births, but the number of these was thought to be negligible, according to the National Center for Health Statistics.

Nonresponse error.

Kindergarten cohort (ECLS-K).). Overall, 880 of the 1,280 eligible schools (69.4 percent weighted) agreed to participate in the fall kindergarten study. Due to the lower-than-expected cooperation rate for public schools in the fall of the base year, 74 additional public schools were included in the sample as substitutes for schools that did not participate. These schools were included in order to meet the target sample sizes for students. Substitute schools are not included in the school response rate calculations.

A nonresponse bias analysis was conducted to determine if substantial bias was introduced due to school nonresponse in the ECLS-K. Five different approaches were used to examine the possibility of bias in the ECLS-K sample. First, weighted and unweighted response rates for schools, children, parents, teachers, and school administrators were examined to see whether there were large response rate differences by characteristics of schools (e.g., urbanicity, region, school size, percent Black, Hispanic, and other race/ethnicity students, grade range) and children (e.g., sex, age, race/ethnicity). Second, estimates based on the ECLS-K respondents were compared to estimates based on the full sample. The distributions of schools by school type, urbanicity, and region, and the distributions of enrollment by kindergarten type (public vs. private), race/ethnicity, urbanicity, region, and eligibility for free and reduced-price lunch were compared for the responding schools and all the schools in the sampling frame. Third, estimates from the ECLS-K were compared with estimates from other data sources (e.g., Current Population Survey, National Household Education Surveys Program, Survey of Income and Program Participation). Fourth, estimates using the ECLS-K unadjusted weights were compared with estimates using the ECLS-K weights adjusted for nonresponse. Large differences in the estimates produced with these two different weights would indicate the potential for bias. Fifth, and last, simulations of nonresponse were conducted. The results of these analyses are summarized in the ECLS-K user's manuals. Findings from these analyses suggest that there is no bias due to school nonresponse.

A total of 940 of the 1,280 originally sampled schools participated during the base year of the study. This translates into a weighted response rate (weighted by the base weight) of 74 percent for the base year of the study. The weighted child base-year survey response rate was 92 percent (i.e., 92 percent of the children were assessed at least once during kindergarten). The weighted parent base-year unit response rate was 89 percent (i.e., a parent interview was completed at least once during kindergarten). Thus, the overall base-year

response rate for children was 68 percent (74 percent of schools x 92 percent of sampled children) and the base-year overall response rate for the parent interview was 66 percent (74 percent of schools x 89 percent of parents of sampled children). About 76 percent of children and 72 percent of parents eligible for the eighth grade data collection (spring 2007) participated.

Birth cohort (ECLS-B). Response rates for all rounds of data collection are determined first and foremost by completion of the corresponding round's parent CAPI instrument. The parent CAPI instrument was chosen as the primary vehicle for determining the overall response rate because there were very few cases (e.g., 0.3 percent at 9 months and 0.06 percent at 2 years) in which other components of the study (e.g., direct child assessments or father questionnaires) were completed but the parent interview was not. All response rates are computed at the child level. In the 9-month data collection, all sampled children were eligible except those children who died before the home visit occurred, children born to mothers younger than 15 years old, children who were adopted before the age of 9 months, and children who were removed from the sample as part of a cost reduction process in February 2002. Response rates for subsequent rounds are conditioned on the completion of a prior round parent interview. For example, the 2-year-round response rate is conditioned on the completion of the 9-month parent interview; all sampled children whose parents completed the 9-month parent component were eligible except those children who had died before the 2-year home visit occurred and children who had moved abroad permanently. For the preschool-year data collection, approximately 9,850 cases with completed 2-year parent interviews, and an additional 50 AI/AN cases with completed 9-month parent interviews, were fielded and considered eligible (approximately 100 children were removed from the sample because they had died or moved abroad permanently). For the kindergarten 2006 collection, there were about 7,000 parent interviews. For the kindergarten 2007 collection, there were about 1,900 parent interviews.

Response rates are also calculated for the other components of the ECLS-B: the child assessments; the resident and nonresident father questionnaires; the care provider interview; the child care observation; the teacher questionnaire; and the school data. Response rates for these other components are conditioned on the completion of the parent interview in all rounds of the

ECLS-B. Only cases with completed parent interviews were assigned weights for the other components of the study.

Table 1. Weighted unit response rates for all children and children sampled in kindergarten, by questionnaire and data collection: Various years 1998–2004

	All children		Children sampled in kindergarten	
Data collection	Child assess-ment	Parent inter-view	Child assess-ment	Parent inter-view
Fall kindergarten	89.9	85.3	†	†
Spring kindergarten	88.0	83.9	†	†
Spring first grade	87.2	83.5	88.0	84.5
Spring third grade	80.1	76.9	80.8	77.8
Spring fifth grade	83.9	88.3	84.7	89.1

† Not applicable.

SOURCE: Tourangeau, K., Burke, J., Le, T., Wan, S., Weant, M., Brown, E., Vaden-Kiernan, N., Rinker, E., Dulaney, R., Ellingsen, K., Barrett, B., Flores-Cervantes, I., Zill, N., Pollack, J., Rock, D., Atkins-Burnett, S., Meisels, S., Bose, J., West, J., Denton, K., Rathbun, A., and Walston, J. (2001). *ECLS-K, Base Year Public-Use Data File, Kindergarten Class of 1998-99: Data Files and Electronic Code Book (Child, Teacher, School Files), and User's Manual* (NCES 2001-029REV). National Center for Education Statistics, U.S. Department of Education. Washington, DC. Tourangeau, K., Burke, J., Le, T., Wan, S., Weant, M., Nord, C., Vaden-Kiernan, N., Bissett, E., Dulaney, R., Fields, A., Byrne, L., Flores-Cervantes, I., Fowler, J., Pollack, J., Rock, D., Atkins-Burnett, S., Meisels, S., Bose, J., West, J., Denton, K., Rathbun, A., and Walston, J. (2002). *User's Manual for the ECLS-K First-Grade Public-Use Data Files and Electronic Codebook* (NCES 2002-135). National Center for Education Statistics, U.S. Department of Education. Washington, DC. Tourangeau, K., Brick, M., Le, T., Wan, S., Weant, M., Nord, C., Vaden-Kiernan, N., Hagedorn, M., Bissett, E., Dulaney, R., Fowler, J., Pollack, J., Rock, D., Weiss, M.J., Atkins-Burnett, S., Hausken, E.G., West, J., Rathbun, A., and Walston, J. (2004). *User's Manual for the ECLS-K Third-Grade Public-Use Data Files and Electronic Codebook* (NCES 2004-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Tourangeau, K., Nord, C., Lê T., Pollack, J.M., and Atkins-Burnett, S. (2006). *Early Childhood Longitudinal Study Kindergarten Class of 1998–99 (ECLS-K), Combined User's Manual for the ECLS-K Fifth-Grade Data Files and Electronic Codebooks* (NCES 2006-032). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

In the 9-month data collection, the weighted completion rate for the parent CAPI instrument was 74.1 percent (table 2). The weighted completion rates for the child assessment, resident father questionnaires,

and nonresident father questionnaires were 95.6, 76.1, and 50.0 percent, respectively.

In the 2-year data collection, the weighted completion rate for the parent CAPI instrument was 93.1 percent. The weighted completion rates for the child assessment, resident father questionnaires, nonresident father questionnaires, child care provider interview, and child care observation (CCO) component were 94.2, 77.7, 39.8, 70.0, and 51.3 percent, respectively. The longitudinal weighted response rates for the parent CAPI instrument, child assessment, and all father questionnaires were 69.0, 65.0, and 48.7 percent, respectively.

In the preschool data collection, the weighted completion rate for the parent CAPI instrument was 91.3 percent. The weighted completion rates for the child assessment, resident father questionnaires, child care provider interview, and CCO component were 98.3, 87.7, 87.4, and 56.8 percent, respectively. The longitudinal weighted response rates for the parent instrument, child assessment, resident father questionnaires, child care provider interview, and CCO component were 63.1, 62.0, 55.3, 55.1, and 35.8, respectively.

In the kindergarten 2006 data collection, the weighted response rate for the parent instrument was 91.8 percent. The weighted unit response rate for the kindergarten 2006 child assessment was 98.6 percent. The weighted unit response rate for the teacher survey for ECLS-B children with a completed parent interview who were enrolled in kindergarten or higher in 2006-07 and were not homeschooled was 75.6 percent; the weighted unit response rate for school data for these same children was 95.9 percent. The overall weighted unit response rate for the parent component after the kindergarten 2006 data collection was 58.0 percent. The longitudinal weighted unit response rates for the parent, child, teacher, and school components after the kindergarten 2006 collection were 58.0, 57.2, 43.8, and 55.6 percent, respectively.

The weighted unit response rate for the kindergarten 2007 parent interview was 92.5 percent. The weighted unit response rate for the kindergarten 2007 child assessment was 99.4 percent. The weighted unit response rate for the teacher survey for ECLS-B children with a completed parent interview who were enrolled in kindergarten or higher in 2007-08 and were not homeschooled was 77.4 percent; the weighted unit response rate for school data for these same children was 96.9 percent. The longitudinal weighted unit response rate for the parent component after the kindergarten 2007 data collection was 53.7 percent.

The overall weighted unit response rates for the child, teacher, and school components after the kindergarten 2007 collection were 53.3, 41.5, and 52.0 percent, respectively.

Table 2. Weighted unit response rates for all children in the ECLS-B, by survey and component: Various years 2001–2007

Component	9-month	2-year	Pre-school	Kinder-garten 2006	Kinder-garten 2007
Parent CAPI	74.1	93.1	91.3	91.8	92.5
Child assessment	95.6	94.2	98.3	98.6	99.4
Resident father	76.1	77.7	87.7	†	†
Nonresident father	50.0	39.8	†	†	†
Child care provider	†	70.0	87.4	†	†
Child care observation	†	51.3	56.8	†	†

† Not applicable.
SOURCE: Denton Flanagan, K., and McPhee, C. (2009). *The Children Born in 2001 at Kindergarten Entry: First Findings From the Kindergarten Data Collections of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)* (NCES 2010-05). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Jacobson Chernoff, J., Flanagan, K. D., McPhee, C., and Park, J. (2007). *Preschool: First Findings From the Preschool Follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)* (NCES 2008-025). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Nord, C., Edwards, B., Andreassen, C., Green, J. L., and Wallner-Allen, K. (2006). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), User's Manual for the ECLS-B Longitudinal 9-Month–2-Year Data File and Electronic Codebook (NCES 2006–046).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Nord, C., Edwards, B., Hilpert, R., Branden, L., Andreassen, C., Elmore, A., Sesay, D., Fletcher, P., Green, J.L., Saunders, R., Dulaney, R., Reaney, L., and Flanagan, K.D. (2004). *User's Manual for the ECLS-B Nine-Month Restricted-Use Data File and Electronic Codebook* (NCES 2004-092). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

An analysis was conducted to assess the potential bias in survey estimates due to unit or item nonresponse for the various components of the survey. This evaluation consisted of several types of comparisons. First, data obtained from children's birth certificates were compared between cases in the sampling frame and sample respondents; data for sample respondents were

weighted first using base weights and then using final weights. These comparisons were made for respondents to the parent CAPI interview, the father questionnaires, the child care provider interview, and the CCO component. In another analysis, birth certificate and survey data were compared between 9-month respondents (using final 9-month weights) and 2-year respondents (using both final 9-month weights and final 2-year weights). These comparisons were done for respondents to the parent CAPI interview, the child assessments, the father questionnaires, and the child care provider interview. The analysis found little or no evidence of potential for bias due to unit nonresponse. Differences between sample respondents and sample frame data were generally small and largely corrected by nonresponse corrections and other adjustments to the base weights. An evaluation comparing the demographic characteristics of respondents and nonrespondents for selected items with less than an 85 percent response rate found no evidence of potential for bias due to item nonresponse. Similar analyses of nonresponse bias were conducted for later rounds of data collection, with no evidence found for bias due to item nonresponse.

Measurement error. In addition to the potential clustering effects related to shared parent SES within schools (described in "Design effects," above), there was a concern in the ECLS-K that the individual mode of administration might inject additional and unwanted variance into both the individual and between-school components of variance in the cognitive scores. Since it is more difficult to standardize test administrations when tests are individually administered, this source of variance could contribute to high design effects if the individual assessors differed systematically in their modes of administration. It was found, however, that the component of variance associated with the individual test administration effect was negligible in all cognitive areas and thus had little or no impact on the design effects.

A potential area for measurement error occurs with the NCATS and Two Bags Task components of the ECLS-B home visit. The parent-child interactions for these two components of the study were videotaped and coded later. The process of coding the tapes, however, is not problem-free. The videotape of the interaction must be of high quality to ensure valid coding. For example, field staff needed to tape the very beginning of the interaction and should not interrupt it. The task of coding is further complicated by the coding staff's experience. Like the ECLS-B home visit field staff, the NCATS and Two Bags Task coders did not, for the most part, possess an extensive background in child development. Training the coding staff to reach 90 percent reliability proved difficult at times and often required additional training.

6. CONTACT INFORMATION

For content information about the ECLS project, contact:

Gail M. Mulligan
Phone: (202) 502-7491
E-mail: gail.mulligan@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Anderson, C., Fletcher, P., and Park, J. (2007). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Psychometric Report for the 2-year Data Collection.* (NCES 2007-084). Washington, DC: National Center for Education Statistics.

Najarian, M., Snow, K., Lennon, J., and Kinsey, S. (2010). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Preschool–Kindergarten 2007 Psychometric Report* (NCES 2010-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Najarian, M. Pollack, J.M., and Sorongon, A.G. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for the Eighth Grade.* (NCES 2009-002). Washington, DC: National Center for Education Statistics.

Pollack, J., Atkins-Burnett, S., Najarian, M., and Rock, D. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for the Fifth Grade.* (NCES 2006-036rev). Washington, DC: National Center for Education Statistics.

Pollack, J., Atkins-Burnett, S., Rock, D., and Weiss, M. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) Psychometric Report for the Third Grade.* (NCES 2005-062). Washington, DC: National Center for Education Statistics.

Rock, D., and Pollack, J. (2002). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) Psychometric Report for Kindergarten through First Grade.* (NCES 2002-05). Washington, DC: National Center for Education Statistics.

Tourangeau, K., Lê, T., Nord, C., and Sorongon, A.G. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) Eighth-Grade Methodology Report.* (NCES 2009-003). Washington, DC: National Center for Education Statistics.

Tourangeau, K., Lê, T., and Nord, C. (2006). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), Fifth Grade Methodology Report.* (NCES 2006-037). Washington, DC: National Center for Education Statistics.

Tourangeau, K., Brick, M., Byrne, L., Lê, T., Nord, C., West, J., and Germino Hausken, E. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) Third Grade Methodology Report.* (NCES 2005-018). Washington, DC: National Center for Education Statistics.

Tourangeau, K., Brick, M., Le, T., Wan, S., Weant, M., Nord, C., Vaden-Kiernan, N., Hagedorn, M., Bissett, E., Dulaney, R., Fowler, J., Pollack, J., Rock, D., Weiss, M.J., Atkins-Burnett, S., Hausken, E.G., West, J., Rathbun, A., and Walston, J. (2004). *User's Manual for the ECLS-K Third-Grade Public-Use Data Files and Electronic Codebook* (NCES 2004-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourangeau, K., Nord., C., Le, T., Wan, S., Bose, J., and West, J. (2002). *User's Guide to the Longitudinal Kindergarten-First Grade Public-Use Data File* (NCES 2002-149). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourangeau, K., Burke, J., Le, T., Wan, S., Weant, M., Nord, C., Vaden-Kiernan, N., Bissett, E., Dulaney, R., Fields, A., Byrne, L., Flores-Cervantes, I., Fowler, J., Pollack, J., Rock, D., Atkins-Burnett, S., Meisels, S., Bose, J., West, J., Denton, K., Rathbun, A., and Walston, J. (2002). *User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Code Book* (NCES 2002-135). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 2: Common Core of Data (CCD)

1. OVERVIEW

The Common Core of Data (CCD) is NCES's primary database on public elementary and secondary education in the United States. Every year the CCD collects information from the universe of state education agencies (SEAs) on all public elementary and secondary schools and education agencies in the United States. The CCD provides descriptive data about staff and students at the school, school district, and state levels. Information about revenues and expenditures is collected at the school district and state levels. Some of the CCD's component surveys date back to the 1930s. The integrated CCD was first implemented in the 1986–87 school year.

SURVEY OF THE UNIVERSE OF ELEMENTARY AND SECONDARY SCHOOLS

CCD collects data through these major components:

- Public Elementary/Secondary School Universe Survey
- Local Education Agency Universe Survey (i.e., school district survey)
- State Nonfiscal Survey of Public Elementary/Secondary Education (i.e., state aggregate nonfiscal survey)
- National Public Education Financial Survey (i.e., state-level financial survey)
- School District Finance Survey
- Teacher Compensation Survey

Purpose

To provide basic statistical information on all children in this country receiving a public education from prekindergarten through grade 12 and information on the funds collected and expended for providing public elementary and secondary education. The specific objectives of the CCD are to (1) provide an official listing of public elementary and secondary schools and education agencies in the nation, which can be used to select samples for other NCES surveys; and (2) provide basic information and descriptive statistics on public elementary and secondary schools and schooling.

Components

There are six components to the CCD: the Public Elementary/Secondary School Universe Survey, Local Education Agency Universe Survey, State Nonfiscal Survey of Public Elementary/Secondary Education, National Public Education Financial Survey (NPEFS), School District Finance Survey, and Teacher Compensation Survey. The CCD surveys consist of data submitted annually to NCES by state education agencies in the 50 states, the District of Columbia, the Bureau of Indian Education (BIE) schools[1], the Department of Defense Dependents Schools, Puerto Rico, and the four outlying areas (American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, and the U.S. Virgin Islands).

Public Elementary/Secondary School Universe Survey. This survey collects information on all public elementary and secondary schools in the United States. (In the 2007–08 school year, there were 101,565 operating and 2,264 nonoperating public elementary and secondary schools.) Data include the school's mailing address, telephone number, operating status, locale (ranging from large city to rural), and type ("regular" or focused on a special area such as vocational education). The survey also collects the student enrollment (membership) for every grade taught in the school; number of students in each of five racial/ethnic groups[2]; number of students eligible for free-lunch programs; and number of classroom teachers, reported as full-time equivalents (FTEs). In the 1998–99 school year,

[1] The BIE assumed administration of these schools from the Bureau of Indian Affairs in 2006.
[2] Student data have been collected in either five or seven racial/ethnic groups since the 2007–08 school year. However, starting in 2011–12, student data will be collected only in seven racial/ethnic groups.

several variables were added: location address (if different from mailing address); Title I, magnet, and charter school status; number of students eligible for reduced-price lunch programs; number of migrant students enrolled the previous year; and enrollment broken out by race and sex within grade.

Local Education Agency Universe Survey. This survey serves as a directory of basic information on local education agencies (LEAs). (In the 2007–08 school year, there were approximately 18,090 LEAs, including 17,941 operating and 149 nonoperating agencies.) It collects the agency's mailing address, telephone number, county location, metropolitan status, and type. The survey includes, for the current year, the total number of students enrolled (membership) in prekindergarten through grade 12; number of ungraded students; number of English language learner (ELL) students served in appropriate programs; and number of instructional, support, and administrative staff. It includes, for the previous year, the number of high school graduates, other completers, and grade 7–12 dropouts. Dropout data were first collected in the 1992–93 CCD, reflecting dropouts for the 1991–92 school year. In 2006–07, the CCD collected both the prior- and current-year number of high school graduates, other completers, and grade 7–12 dropouts. Since 2007–08, however, only current-year data on high school completers and dropouts have been collected. Also, since 2007-08, the high school dropout and completion data have been separated from the LEA universe survey data and released as standalone data files.

State Nonfiscal Survey of Public Elementary/Secondary Education. This survey collects information on all students and staff aggregated to the state level, including number of students by grade level; counts of FTE staff by major employment category; and high school completers by race/ethnicity. Since 2007-08, data on student enrollment and staffing are for the current school year. Through school year 2005–06, data on high school completers and dropouts were collected for the previous year. The collection cycle for school year 2006–07 was a transition year when both prior- and current-year data on high school completers and dropouts were collected.

National Public Education Financial Survey (NPEFS). This survey collects detailed finance data at the state level, including average daily attendance, school district revenues by source (local, state, federal), and expenditures by function (instruction, support services, and noninstruction) and object (salaries, supplies, etc.). It also reports capital outlay and debt service expenditures. Revenues and expenditures are audited after the close of the fiscal year and are then submitted to NCES by each state education agency.

The NPEFS underwent a major revision in fiscal year (FY) 1989, acquiring its present name in that year and greatly increasing the number of data items collected. Since that year, additional items have been added to and deleted from the survey. In the FY 89 data collection, NCES also began providing "crosswalk" software to assist states in their reporting and to improve the comparability of data across states. This software converts a state's existing accounting reports to uniform federal standards, as described in the NCES accounting handbook (National Forum on Education Statistics 2003). The most recent change in the NPEFS is the addition of teacher salary expenditures broken out by program (regular, special education, vocational, and other education program), as well as the addition of textbook expenditures. Data on expenditures from the America Reinvestment and Recovery Act will be collected and reported separately for fiscal years 2009 through 2011.

School District Finance Survey. This survey collects detailed data by school district, including revenues by source, expenditures by function and subfunction, and enrollment. These data are collected by the Governments Division of the U.S. Census Bureau and are released as the Annual Survey of Local Government Finances (F-33). Before FY 95, data were collected from all districts in decennial census years (e.g., 1990) and years ending in 2 and 7, and from a large sample in other years. The F-33 was first conducted in FY 80. Beginning with FY 95, detailed fiscal data on revenues and expenditures have been collected for all school districts providing public education to students in prekindergarten through grade 12. These data can be linked to the nonfiscal data collected in the Local Education Agency Universe Survey. Student counts and amounts of debt at the beginning and end of the fiscal year are also provided. NCES began to substantially support the F-33 in FY 92.

In FY 97, two variables, Payments to Private Schools and Payments to Public Charter Schools, were added. In FY 1998, two variables that describe the nature of school districts and their relation to other surveys and data files were added: AGCHRT and CENFILE. AGCHRT identifies school districts with charter schools, and CENFILE identifies those districts that are available in the Census Bureau's version of the F-33 school district file. Similar to changes in the NPEFS, teacher salary and textbook exhibit items were added to the F-33 beginning with the FY 04 collection. Special exhibit items are separate data items that are included in, but do not summarize to, other data items. Starting with the FY 05 collection, the data item Federal Revenue—Bilingual Education (B11) was moved from the "federal revenue direct" section to the "federal revenue through the state" section. This change was made as a result of changes in the allocation of

bilingual education funds by the U.S. Department of Education. In the FY 06 collection, four new local revenue items were added: rents and royalties, sale of property, fines and forfeits, and private contributions. Data on expenditures from the America Reinvestment and Recovery Act will be collected and reported separately for fiscal years 2009 through 2011.

Teacher Compensation Survey. This survey collects total compensation, teacher status, and demographic data about individual teachers from multiple states. In 2007, NCES launched the pilot Teacher Compensation Survey (TCS) data collection, with seven states volunteering to provide administrative records for school year (SY) 2005-06. The TCS expanded to 17 states reporting SY 2006-07 and SY 2007-08 data. Twenty-three states are currently participating and up to 35 states will volunteer to participate in the TCS from 2010 to 2013. The TCS file can be merged with the CCD Public Elementary/Secondary School Universe Survey file. Unique ID numbers are used to track teachers within states over time. The data are released as a restricted-use file, available to researchers with an IES data license. The data items on the restricted-use file include: Teacher ID, NCES School ID, FTE, base salary, total salary, employee benefits, years of teaching experience, highest degree earned, race, age, and teacher status codes. Teachers at more than one school will have a record for each school they teach in, and the FTE and salary values are for the teacher at that school only. Summary descriptive statistics are released in public use files. The public use files include teachers' mean base salary, level of education, and mean base salary by varying levels of experience at the school and LEA level.

Periodicity

Annual. Some of the component surveys were initiated during the 1930s. In its integrated form, the CCD was introduced in the 1986–87 school year.

2. USES OF DATA

The CCD collects three categories of information: (1) general descriptive information on schools and school districts, including name, address, phone number, and type of locale; (2) data on students and staff, including demographic characteristics (e.g., race/ethnicity); and (3) fiscal data covering revenues and current expenditures. The datasets within the CCD can be used separately or jointly to provide information on many topics related to education. The ease of linking CCD data with other datasets makes the CCD an even more valuable resource.

Not only is the CCD a source of data that can be used to demonstrate relationships between different school, district, and state characteristics, it can also provide a historical record of schools or agencies of interest. This information can shed light on how and why education in the United States is changing. The types of schools or districts that have changed the most with respect to a measured characteristic (e.g., proportion of Hispanic students) can be identified, and the reasons for these changes can be independently investigated. Similarly, the impact of state and local education policies and practices can be assessed through an examination of changes in school and district characteristics. For example, districts that have shown substantial improvement in their racial balance or interracial exposure indices can be identified. The policies and practices employed by these districts can then be examined. By identifying the presence of significant changes and where these changes are occurring, CCD data can help policymakers and practitioners better target their efforts and help researchers develop more sharply focused hypotheses for investigating key education issues.

3. KEY CONCEPTS

The concepts described below pertain to the levels of data collection (school, agency, state) and school locale in the CCD. For a comprehensive list of CCD terms and definitions, refer to the glossaries in various CCD annual publications (such as CCD files and documentation, First Look reports, and technical user guides) available on the Internet (http://nces.ed.gov/ccd/ccd.publications).

Local Education Agency. An LEA has administrative responsibility for providing instruction or specialized services to one or more elementary or secondary schools. Most LEAs are *regular school districts* that are locally administered and directly responsible for educating children. Others are *supervisory unions* (which provide administrative systems for the smaller regular districts with which they are associated); *regional education service agencies* (which offer research, data processing, special education or vocational program management, and other services to a number of client school districts); *state-operated school districts* (e.g., for the deaf and blind); *federally operated school districts* (e.g., operated by the Bureau of Indian Education); and *other agencies* not meeting the definitions of the preceding categories (e.g., operated by a Department of Corrections). Since school year 2007–08, a charter agency type code has been used to differentiate charter agencies from other types of agencies.

Public Elementary/Secondary School. An institution that is linked with an education agency, serves students, and has an administrator. It is possible for

more than one CCD-defined school to exist at a single location (e.g., an elementary and secondary school sharing a building, each with its own principal). One school may also be spread across several locations (e.g., a multiple "storefront" learning center managed by a single administrator).

The CCD classifies schools by type. *Regular schools* provide instruction leading ultimately toward a standard high school diploma; they may also offer a range of specialized services. *Special education* and *vocational schools* have the provision of specialized services as their primary purpose. *Other alternative schools* focus on an instructional area not covered by the first three types (e.g., developing basic language and numeracy skills of adolescents at risk of dropping out of school).

Some schools do not report any students in *membership* (i.e., enrolled on the official CCD reporting day of October 1). This occurs when students are enrolled in more than one school but are reported for only one. For example, students whose instruction is divided between a regular and a vocational school may be reported only in membership for the regular school. In other cases, a school may send the students for which it is responsible to another school for their education—a situation most likely in a small community that does not have sufficient students to warrant keeping a school open every year.

School Locale. Beginning with the 2006-07 CCD files, the locale code methodology was changed from a 1-digit code based on metropolitan statistical areas to a 2-digit code based on urban clusters. American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, the U.S. Virgin Islands, and the Department of Defense Dependents Schools (overseas) were not assigned a locale code because the geographic and governmental structures of these entities do not fit the definitional scheme used to derive the code. There are eight metro-centric locale codes.

The new "urban-centric" locale codes are assigned through a methodology developed by the U.S. Census Bureau's Population Division in 2005. The urban-centric locale codes apply current geographic concepts to the NCES locale codes used from 1986 through the present. The new urban-centric methodology supplements, and will eventually replace, the older locale code methodology. American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, the U.S. Virgin Islands, and the Department of Defense Dependents Schools (overseas) were not assigned a locale code because the geographic and governmental structures of these entities do not fit the definitional scheme used to derive the code. The Department of Defense Dependents Schools (domestic) were not assigned locale codes because it is not legal to do so. The new system has 12 urban-centric locale codes.

4. SURVEY DESIGN

Target Population

All public elementary and secondary schools, LEAs, and SEAs throughout the United States, including the District of Columbia, the overseas Department of Defense Dependents Schools, BIE schools, Puerto Rico, and the four outlying areas.

Sample Design

The CCD collects information from the universe of state-level education agencies, except for the Teacher Compensation Survey. The Teacher Compensation Survey is a new survey, and states are participating in it when they are able to report the requested data.

Data Collection and Processing

Through the 2005–06 collection, CCD data were voluntarily obtained from administrative records collected and edited by SEAs during their regular state reporting cycle. In 2006–07, CCD nonfiscal data reporting became mandatory for SEAs. In 2007–08, reporting CCD nonfiscal data to EDFacts, a new data collection system, became mandatory for SEAs.

Reference dates. Most data for the nonfiscal surveys are collected for a particular school year (September through August). The official reference date is October 1st or the closest school day to October 1st. Special education, free-lunch eligibility, and racial/ethnic counts may be taken on December 1st or the closest school day to that date. Student and teacher data are reported for the current school year, whereas through 2005–06, data for high school graduates, other completers, and dropouts reflected the previous year. Fiscal data are for the previous fiscal year; thus, FY 98 data represent the 1997–98 school year.

Data collection. The ways in which CCD data are collected have evolved with the advancement of technology. In the early days of the collection, survey instruments were usually distributed to the states in January. Starting in the 2001–02 collection, downloadable PC software was used. In 2004–05, a web-based data collection application was developed and put into use. A state CCD coordinator, appointed by the Chief State School Officer, is responsible for overseeing the completion of the surveys (often, different coordinators are responsible for the fiscal and the nonfiscal surveys). To assure comparable data across states, NCES provides the CCD coordinator with a set of standard critical definitions for all survey items. In addition, data conferences and training

sessions are held at least yearly. The state's data plan identifies any definitional differences between the state's recordkeeping and the CCD's collection as well as any adjustments made by the state to achieve comparability. Counts across CCD surveys may not be identical, but differences should be consistent and the state is asked to describe the reason for any discrepancy.

NCES provides the state with general information collected during the previous survey on each district and school (e.g., name, address, phone number, locale code, and type of school/district). This information must be verified as correct by the CCD coordinator or recoded with the correct information. The coordinator must also assign appropriate identification codes to new schools and agencies and update the operational status codes for schools and agencies that have closed.

Beginning with the 2005-06 school year, the CCD nonfiscal data have been collected through the U.S. Department of Education's Education Data Exchange Network (EDEN). States report data to EDEN through multiple file groups that fall into various reporting schedule throughout the year. Although states may report data outside the collection period and may revise their reported data at any time in EDEN, NCES extracts the data files from EDEN on the cutoff dates of data submission. The data resubmitted by states after the files were extracted may or may not be included in the CCD final release file.

Data for the CCD fiscal surveys and the TCS are collected by the U.S. Census Bureau. The data are compiled into prescribed formats and submitted by the SEAs. The closing date for the current year's data is the Tuesday following Labor Day. Corrections to submitted fiscal data are accepted until October 1st, however, only corrections that lower a state's current expenditure per pupil are accepted after the "Labor Tuesday" deadline for use in the formula for allocating Title I and other Department of Education funding to state and local school systems.

Editing. Completed surveys undergo comprehensive editing by NCES and the states. Where data are determined to be inconsistent, missing, or out of range, NCES contacts the SEAs for verification. States are given the edit software or are provided with access to the designated website that NCES uses to review data. They are also asked to confirm prepared summaries of the collected information. At this time, the states may revise data collected in the previous survey cycle. NCES examines the data from the 120 largest school districts on a record-by-record basis, setting up fail-safe edit checks to catch unexplained anomalies. In addition, records are processed through a post-edit check to replace blanks and nonmeaningful zeroes with meaningful responses. After editing, final adjustments for missing data are performed.

Estimation Methods

NCES estimates missing values to improve data comparability across states. Only state-level data are estimated on a regular basis. Missing values in the Public School Universe and Local Agency Universe Surveys are generally left as missing, with a few exceptions. No imputations or adjustments are conducted for state-level data on high school graduates, other high school completer categories, or race/ethnicity.

There are two basic estimation methods: imputation and adjustment. *Imputation* is performed when the missing value for a data item is not reported at all indicating that subtotals and totals containing the category are underreported. Imputation assigns a value to the missing item, and the subtotals and totals containing this item are increased by the amount of the imputation. *Adjustment* corrects a situation in which a value reported for one item contains a value for one or more additional items not reported elsewhere. The original value is reduced by an appropriate amount, which is distributed to the items missing a value. All totals and subtotals are then recalculated. If it is not possible to impute or adjust for a missing value, the item is set to -1 and is counted as "missing."

Every cell in the data file has a companion cell with a flag indicating whether the data contents were reported by the state (R) or placed there by NCES using one of several methodologies: adjustment (A); imputation based on the prior year's data (P); imputation based on a method other than the prior year's data (I); totaling based on the sum of internal or external detail (T); or combining with data provided elsewhere by the state (C).

Estimating state-level nonfiscal data. NCES imputes and adjusts some reported values for student and staff counts at the state level (including the District of Columbia). Imputations for prekindergarten students are performed first, followed by staff imputations and then other adjustments. No imputations or adjustments are made to racial/ethnic data.

Estimating state-level fiscal data. NCES also imputes and adjusts revenue and expenditure data. The federal standard (see National Forum on Education Statistics 2003) is used in the adjustments to distribute expenditure and revenue data. Adjustments are also used to distribute direct state support expenditures to specific objects and functions. In some cases, local revenues from student activities and food services are imputed.

Future Plans

Because it is an ongoing annual survey, the CCD engages in continuous planning with its data users and providers.

5. DATA QUALITY AND COMPARABILITY

The data in the CCD are obtained from the universe of SEAs, which are provided with a common set of definitions for all data items requested. In addition, for the CCD fiscal surveys, NCES provides crosswalk software that converts a state's existing accounting reports to the federal standard, as indicated in *Financial Accounting for Local and State School Systems, 2003 Edition* (National Forum on Education Statistics 2003). This ensures the most comparable and comprehensive information possible across states. As with any survey, however, there are possible sources of error, as described below.

Sampling Error

Because the CCD is a universe survey, its data are not subject to sampling errors.

Nonsampling Error

Coverage error. An NCES report by Owens and Bose (1997), found that overall coverage in the 1994–95 Local Education Agency Universe Survey was 96.2 percent of that in state education directories. "Regular" agencies—those traditionally responsible for providing public education—had almost total coverage. Most coverage discrepancies were attributed to nontraditional agencies that provide special education, vocational education, and other services.

Nonresponse error

Unit nonresponse. The unit of response in the CCD is the SEA. Under current NCES standards, the regular components of the CCD are likely to receive at least partial information from every state, resulting in a 100 percent unit response rate.

Item nonresponse. Any data item missing for one school district is generally missing for other districts in the same state. The following items have higher than normal nonresponse: free-lunch-eligible students by school; nontraditional agencies; and dropouts. Some states assign all ungraded students to one grade and therefore do not report any ungraded students.

Several items have shown marked improvement in response during recent years. Student enrollment was only reported for 80 percent of the districts in 1986–87, but is now available for almost 100 percent. Reports of student race/ethnicity at the school level has increased from 63 percent in the 1987–88 school year (when first requested) to nearly 100 percent today.

Measurement error. Measurement error typically results from varying interpretations of NCES definitions, differing record keeping systems in the states, and failures to distinguish between zero, missing, and inapplicable in the reporting of data. NCES attempts to minimize these errors by working closely with the state CCD coordinators.

Definitional differences. Although states follow a common set of definitions in their CCD reports, the differences in how states organize education lead to some limitations in the reporting of data, particularly regarding dropouts. CCD definitions appear to be less problematic in the NPEFS, although data on average daily attendance in this survey are not comparable across states. States provide figures for average daily attendance in accordance with state law; NCES provides a definition for states to use in the absence of state law. Because of this lack of comparability, student membership counts from the State Nonfiscal Survey are used as the official state counts.

Because not all states follow the CCD dropout definition and reporting specifications, dropout counts cannot be compared accurately across states. For states that do not comply with the CCD definition, the dropout count is blanked out in the database and considered missing. Currently, there is considerable variation across local, state, and federal data collections on how to define dropouts. The CCD's definition differs from that in other data sources, including the High School and Beyond Study, the National Education Longitudinal Study of 1988, and the Current Population Survey (CPS), conducted by the Census Bureau. Although the collection of dropout information in the CCD is designed to be consistent with procedures in the CPS, differences remain. CCD dropout data are obtained from state administrative records (whereas the CPS obtains this information from a household survey). The CCD includes dropouts in grades 7 through 12 (whereas the CPS includes only grades 10 through 12).

States also vary in the kinds of high school completion credentials on which they collect data. Some states issue a single diploma regardless of the student's course of study. Others award a range of different credentials depending upon whether the student completed the regular curriculum or addressed an individualized set of education goals. Unreported information is shown as missing in CCD data files and published tables unless it is possible to impute or adjust a value (see "Estimation Methods" in section 4 above).

Changes in state reporting practices. The basic characteristics of a school or district do not change frequently. However, a minor change in local or statewide reporting practices (such as two or three coordinators instructing schools to review all of their general information) can have a large impact on the reliability and validity of CCD items. In the 1990–91 school year, a significant proportion (7 percent) of schools, primarily in three states, reported a change in locale code from the prior survey. While this undoubtedly provided better information on school locales in these states, data became less comparable across years. Such changes are rare, however, and tend to be clustered by state and year.

Data Comparability

Most CCD items can be used to assess changes over time by state, district, and school. However, checks of the prevalence and patterns of nonresponse should be performed to assess the feasibility of any analysis. There may also be discontinuities in the data resulting from the introduction of new survey items, changes in state reporting practices, etc., and there may be inconsistencies across reporting levels in the numbers for the same data element (e.g., number of students).

Content changes. As new items are added to the CCD, NCES encourages states to incorporate into their own survey systems the items they do not already collect so that these data will be available in future rounds of the CCD. Over time, this has resulted in fewer missing data cells in each state's response, thus reducing the need to impute data. Users should keep in mind, however, that while the restructuring of data collection systems can produce more complete and valid data, it can also make data less comparable over time. For example, prior to FY 89, public revenues were aggregated into four categories and expenditures into three functions. Because these broad categories did not provide policymakers with sufficient detail to understand changes in the fiscal conditions of states, the survey was expanded in 1990 to collect detailed data on all public revenues and expenditures within states for regular education in prekindergarten through grade 12.

Comparisons within the CCD. A major goal of the CCD is to provide comparable information across all surveys. The surveys are designed so that the schools in the Public School Universe survey are reflected in the Local Agency Universe survey and so that the data from these surveys are reflected in the State Nonfiscal survey. While counts may not always be equal across reporting levels or even within the same level, differences should be consistent and explainable. For example, counts of students by race/ethnicity in the Public School Universe survey may not always be comparable to student counts by grade because these counts may be taken at different times.

For the most part, the total number of students in a regular district is close to the aggregated number of students in all of the district's schools. Since 1990, there has typically been agreement between these counts in at least 85 percent of the districts. Membership numbers in the Public School Universe and Local Agency Universe surveys may legitimately differ if (1) there are students served by the district but not accounted to any school (e.g., hospitalized or homebound students); or (2) there are schools operated by the state Board of Education rather than by a local agency. To avoid confusion, NCES publishes the numbers of students and staff from the State Nonfiscal Survey as the official counts for each state.

Teacher counts may also vary across reporting levels. For example, FTE teacher counts are rounded to the nearest hundredth in the Public School Universe survey, but to the nearest whole number in the State Nonfiscal Survey.

6. CONTACT INFORMATION

For content information on the CCD, contact the following individuals:

Program Director:
Marie Stetser
Phone: 202-502-7356
E-mail: marie.stetser@ed.gov

Nonfiscal Surveys:
Robert Stillwell
Phone: (202) 219-7044
E-mail: robert.stillwell@ed.gov

Nonfiscal data collection and related publications:
Patrick Keaton
Phone: (202) 502-7386
E-mail: patrick.keaton@ed.gov

Fiscal data collection and related publications:
Frank Johnson
Phone: (202) 502-7362
E-mail: frank.johnson@ed.gov

Teacher Compensation Survey:
Stephen Cornman
Phone: (202) 502-7338
E-mail: stephen.cornman@ed.gov

Mailing Address for All Contacts:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education

1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Sable, J., and Plotts, C. (2009). *Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2007–08* (NCES 2010-302). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Sable, J., Plotts, C., and Tang, J. (2009). *Documentation to the NCES Common Core of Data Local Education Agency Universe Survey: School Year 2007–08* (NCES 2010-303). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Hamann, T. (2000). *Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary School Universe Survey* (NCES Working Paper 2000-12). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Hoffman, L. (1997). *Customer Service Survey: Common Core of Data Coordinators* (NCES Working Paper 97-15). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Owens, S. and Bose, J. (1997). *Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary Education Agency Universe Survey* (NCES 97-505). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Parrish, T.B., Matsumoto, C.S., and Fowler, W.J. (1995). *Disparities in Public School District Spending 1989–90: A Multivariate, Student-Weighted Analysis, Adjusted for Differences in Geographic Cost of Living and Student Need* (NCES 95-300). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Survey Design

Hamann, T.A. (1999). *Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle* (NCES Working Paper 1999-03). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 3: Private School Universe Survey (PSS)

1. OVERVIEW

BIENNIAL SURVEY OF THE UNIVERSE OF PRIVATE SCHOOLS

PSS collects data on:

- Student enrollment
- Teaching staff
- High school graduates
- School religious affiliation

In recognition of the importance of private education, NCES has made the collection of data on private elementary and secondary schools a priority. In 1988, NCES introduced a proposal to develop a private school data collection system that would improve on the irregular collection of private school information dating back to 1890. Since 1989, the U.S. Census Bureau has conducted the biennial Private School Universe Survey (PSS) for NCES. The PSS collects information comparable to that collected on public schools in the Common Core of Data (CCD) (see chapter 2). PSS data are complemented by the more in-depth information collected in the private school sample surveys that are part of the Schools and Staffing Survey (SASS) (see chapter 4). The next PSS data collection will take place during the 2011-12 school year. The next SASS is also planned for the 2011–12 school year.

Purpose

To (1) build an accurate and complete universe of private schools to serve as a sampling frame for NCES surveys of private schools; and (2) generate biennial data on the total number of private schools, teachers, and students.

Components

The PSS consists of a single survey that is completed by administrative personnel in private schools. An early estimates survey designed to allow early reporting of key statistics was discontinued after the 1992–93 school year.

Private School Universe Survey. This survey collects data on private elementary and secondary schools, including religious orientation, level of school, length of school year, length of school day, total enrollment (K–12), race/ethnicity of students, number of high school graduates, number of teachers employed, program emphasis, and existence and type of kindergarten program.

Periodicity

Biennial. The next PSS will be administered in 2011-12 and every 2 years thereafter. Earlier surveys were conducted in 1989–90, 1991–92, 1993–94, 1995–96, 1997–98, 1999–2000, 2001–02, 2003–04, 2005–06, and 2009-10.

2. USES OF DATA

The PSS produces private school data similar to that produced for public schools in the CCD. Profiles of private education providers can be developed from PSS data to address a variety of policy- and research-relevant issues, including the growth of religiously affiliated schools, the number of private high school graduates, the

length of the school year for various private schools, and the number of private school students and teachers.

3. KEY CONCEPTS

Some key concepts related to the PSS are described below.

Private School. A school that is not supported primarily by public funds. It must provide classroom instruction for one or more of grades K–12 (or comparable ungraded levels) and have one or more teachers. Organizations or institutions that provide support for home schooling but do not offer classroom instruction for students are not included. Private schools are assigned to one of three major categories and, within each major category, to one of three subcategories:

- *Catholic:* parochial, diocesan, private;
- *Other religious:* affiliated with a conservative Christian school association, affiliated with a national denomination, unaffiliated; and
- *Nonsectarian:* regular program emphasis, special program emphasis, special education.

Schools with kindergarten, but no grade higher than kindergarten, are referred to as *kindergarten-terminal (K-terminal)* schools; these schools were first included in the 1995–96 PSS. Schools meeting the pre-1995 definition of a private school (i.e., including any of grades 1–12) are referred to as *traditional* schools.

Elementary School. A school with one or more of grades K–6 and no grade higher than grade 8. For example, schools with grades K–6, 1–3, or 6–8 are classified as elementary schools.

Secondary School. A school with one or more of grades 7–12 and no grade lower than grade 7. For example, schools with grades 9–12, 7–8, 10–12, or 7–9 are classified as secondary schools.

Combined School. A school with one or more of grades K–6 and one or more of grades 9–12. For example, schools with grades K–12, 6–12, 6–9, or 1–12 are classified as combined schools. Schools in which all students are ungraded (i.e., not classified by standard grade levels) are also classified as combined.

Teacher. Any full- or part-time teacher whose school reports that his or her assignment is teaching in any of grades K–12.

4. SURVEY DESIGN

Target Population

All private schools in the United States that meet the NCES definition. The PSS universe consists of a diverse population of schools. It includes both schools with a religious orientation (e.g., Catholic, Lutheran, or Jewish) and nonsectarian schools with programs ranging from regular to special emphasis and special education.

Sample Design

NCES uses a dual-frame approach for building its private school universe. The primary source of the PSS universe is a *list frame* containing most private schools in the country. The list frame is supplemented by an *area frame*, which contains additional schools identified during a search of randomly selected geographic areas around the country. The two frames are used together to estimate the population of private schools in the United States. Since documentation for the 2009-10 PSS has not been completed, these descriptions are for the 2007-08 PSS.

List frame. In an effort to ensure a complete population list of all private elementary and secondary schools in the United States, NCES updates the list frame every 2 years in preparation for the next PSS administration. The list frame was initially developed for the 1989-90 survey. The list is updated periodically by matching it with lists provided by nationwide private school associations, state departments of education, and other national private school guides and sources.

The basis of the current survey's list frame is the previous PSS. In order to expand coverage to include private schools founded since the previous survey, NCES requests lists of schools from the 50 states and the District of Columbia in advance of each survey administration. Requests are made to state education departments, as well as to other departments, such as health or recreation. NCES also collects membership lists from about 29 private school associations and religious denominations. Schools on the state and association lists are compared to the base list, and any school not matching a school on the base list is added to the universe list.

Prior to the 1995–96 survey, only schools that included at least one of grades 1–12 were included in the PSS (now referred to as *traditional schools*). As of 1995–96, the PSS has also collected data from K-terminal schools. NCES also removed from the PSS eligibility criteria the requirements that a school have 160 days in the school year and 4 hours per day during which classes are conducted.

In 2007, a separate list-building operation (Early Childhood Operation) was conducted to identify K-terminal schools. Requests for lists of programs that might include a kindergarten were made to sources, other than state departments of education, in all 50 states and the District of Columbia, including state departments of health or recreation; state child care licensing agencies; and child care referral agencies. In 2007, some 24 of these early childhood lists were received, and 19 were processed (due to resource constraints, not all of the lists were processed).

Schools on private school association membership lists, the state lists, and the early childhood lists were compared to the base list, and any school that did not match a school on the base list was added to the universe list. Additionally, questionnaires were sent out to programs identified in the 2005–06 PSS as prekindergarten only. This procedure was done in case any of these programs included at least a kindergarten in the 2007–08 school year. A total of 37,275 schools (unweighted) were included in the 2007–08 list frame.

Area frame. The list frame is supplemented by an area frame, which contains additional private schools identified during a search of telephone books and other sources in randomly selected geographic areas around the country. The area frame search is conducted by the Bureau of the Census. Each area's list is created from a set of predetermined sources within that area and then matched against the updated list frame universe to identify schools missing from the updated list frame.

The United States is divided into 2,062 primary sampling units (PSUs), each consisting of a single county, independent city, or cluster of geographically contiguous areas. The eight PSUs with the highest private school enrollment in the 2000 census populations greater than 1.7 million were selected with certainty for the private school survey. In addition to these certainty PSUs, the area frame consists of two sets of sample PSUs: (1) a 50 percent subsample (overlap) of the area frame sample PSUs from the previous PSS, to maintain a reasonable level of reliability in estimates of change, and (2) a sample of PSUs selected independently from the previous PSS sample (nonoverlap PSUs). A minimum of two nonoverlap PSUs are allocated to each of the 16 strata, which are defined by (1) four Census regions (Northeast, Midwest, South, or West); (2) metro/nonmetro status (two levels); and (3) whether the PSU's percentage of private school enrollment exceeds the median percentage of private enrollment of the other PSUs in the census region/metro status strata (two levels). Within a stratum, the sample PSUs are selected with probability proportional to the square root of the population in each of the PSUs.

A total of 124 distinct PSUs (162 counties) were in the 2007–08 PSS area frame sample. Within each of these PSUs, the Census Bureau attempted to find all eligible private schools. A block-by-block listing of all private schools in a sample of PSUs was not attempted. Rather, regional office field staff created the frame by using such sources as the yellow pages, local Catholic dioceses, religious institutions local education agencies, and local government offices. Once the area search lists were constructed, they were matched with the NCES private school universe list. Schools that did match the universe list were deleted from the area frame. A total of 1,872 schools (unweighted) were added to the universe from the area frame.

Due to differences in methodology and definition, the results of the 1993–94 and subsequent area search frames are not strictly comparable to the results of earlier years. Prior to 1993, an initial eligibility screening was performed by telephone for area frame schools before the questionnaire was mailed out. Ineligible schools were declared out of scope at that time, and eligible schools were either interviewed by telephone or sent a questionnaire. In the 1993–94 PSS, screener questions were added to the survey instrument to determine eligibility. Ineligible schools were not eliminated until the questionnaires were returned. In the 1995–96 PSS, all area frame schools were placed in the telephone follow-up phase of the PSS, and ineligible schools were again eliminated based on responses to screener questions.

Data Collection and Processing

The data collection phase consists of (1) a mailout/ mailback stage; and (2) a telephone follow-up stage. The U.S. Census Bureau is the collection agent.

Reference dates. The official reference date for reporting PSS information is October 1.

Data collection. In October of the survey year, the Census Bureau mails PSS questionnaires to the private schools. (Data collection for the 2007–08 PSS coincided with the data collection phase of the private school component of the 2007–08 SASS: the private schools selected for SASS were excluded from the PSS, and the schools selected for SASS received a SASS private school questionnaire only, while the remaining private schools were sent a PSS questionnaire. The PSS questionnaire used the same wording as the SASS questionnaire, but contained only a subset of the SASS questionnaire items. After data collection, the data for the SASS cases were merged into the PSS universe.) If no response is received within a month, a second questionnaire is mailed. Reminder postcards are sent 1 week after each questionnaire mailout. Three to 4 months after the initial mailout, the Census Bureau begins telephone follow-up of schools that have not responded to either

mailout; the schools from the area frame operation are added at this time. Interviewing takes place at the Census Bureau's computer-assisted telephone interviewing (CATI) facilities. For schools that cannot be contacted by telephone, additional follow-up is conducted in the Census Bureau's regional offices.

Editing. Most of the mailback questionnaires are scanned; those that must be keyed are 100 percent key-verified. For data collected during the telephone follow-up phase, preliminary quality assurance and editing checks take place at the time of the interview. The data collection instrument is designed to alert interviewers to inconsistencies reported by the respondent so that any necessary corrections can be made at this time. Data from the CATI facilities are transmitted to Census headquarters for further processing where they undergo extensive editing, including:

- range checks to eliminate out-of-range entries;
- consistency edits to compare data in different fields for consistency;
- blanking edits to verify that skip patterns on the questionnaire were followed; and
- interview status recodes (ISRs), performed prior to the weighting process, to assign the final interview status to the records (i.e., interview, noninterview, or out-of-scope).

Estimation Methods

Weighting adjusts the number of schools in the area frame sample up to a fully representative number of schools missing from the list frame and adjusts the survey data from both the area and list components for school nonresponse. Imputation is used to compensate for item nonresponse.

Weighting. PSS data from the area frame component are weighted to reflect the sampling rates (probability of selection) in the PSUs. Survey data from both the list and area frame components are adjusted for school nonresponse. This represents a departure from procedures used in the 1989–90 survey, which adjusted for total nonresponse (i.e., school nonresponse) and for partial nonresponse associated with four specific PSS data elements. Since 1991, only one weight has been required, due to a newly developed and complex imputation process used to compensate for item nonresponse. When estimates are produced for schools and other data elements, the same PSS school weight should be used. A brief description of the components comprising the PSS weight follows:

W_i, the PSS weight for all data items for the i^{th} school, is

$$W_i = BW_i \times NR_c$$

where BW_i is the ***base weight***, or the inverse of the selection probability for school i (BW_i = 1 for list frame schools; BW_i = the inverse of the PSU probability of selection for area frame schools), and

NR_c is the ***nonresponse adjustment factor***, or weighted ratio of the sum of the in-scope schools to the sum of the in-scope responding schools in cell c, using BW_i as the weight.

The cells used to compute the nonresponse adjustment are defined differently for list-frame and area-frame schools. In 2007-08 PSS, for schools in the list frame, the cells were defined by affiliation, urbanicity type, grade level, region, and enrollment. The nonresponse adjustment cells for area frame schools were defined by certainty/noncertainty PSU status, three-level typology (Catholic, Other religious, Nonsectarian), and grade level.

If the number of schools in a cell was less than 15 or the nonresponse adjustment factor was greater than 1.5, then that cell was collapsed into a similar cell. The cells for traditional schools from the list frame were collapsed within enrollment category, urbanicity type, grade level, and census region. Cells for K-terminal schools from the list frame were collapsed within enrollment category, urbanicity type, region (if applicable), and affiliation. Cells for traditional schools from the area frame were collapsed within grade level and then within three-level typology. Cells for K-terminal schools from the area frame were collapsed within three-level typology.

Imputation. Since the 1991–92 PSS, imputation has been used to compensate for item nonresponse in records classified as interviews (i.e., required items are completed). All items that are missing data are imputed. The first survey, the 1989–90 PSS, used weighting adjustments for both interviews and noninterviews.

Imputation occurs in two stages. The first-stage (internal) process uses data from other items for the same school in the current PSS and data from the previous PSS. If an item cannot be imputed during the first-stage process, it is imputed during the second stage. The second-stage (donor) process uses a hot-deck imputation methodology that extracts data from the record for a reporting school (donor) similar to the nonrespondent school. All records (donors and nonrespondents) in the file are sorted by variables that describe certain characteristics of the schools, such as

school type, affiliation, school level, enrollment, and urbanicity.

For a few items, entries are clerically imputed. The data record, sample file record, and the questionnaire are reviewed, and an entry consistent with the information from those sources is imputed. This procedure is used when: (1) no suitable donor is found, (2) the computer method produces an imputed entry that is unacceptable, and (3) the nature of the item requires an actual review of the data rather than a computer-generated value.

Recent Changes

Several changes to the questionnaire have been introduced in the previous PSS cycles. In the 1993–94 PSS, three major revisions were made. First, a new design was implemented to facilitate respondent reporting by clearly indicating skip patterns through the use of arrows as well as words and by minimizing the number of questions asked on each page. Second, content on prekindergarten programs was expanded to collect the type of prekindergarten program in addition to the prekindergarten student and teacher counts requested in earlier surveys (these data were collected as a part of a separate Census Bureau initiative and are not included in PSS reports). Third, data on the racial/ethnic makeup of the school's student body were collected for the first time.

Modifications made to the 1995–96 PSS included adding nursery and prekindergarten, transitional kindergarten, and transitional first-grade enrollment counts to the enrollment item. Questions on the length of the school day and number of days per week for kindergarten, transitional kindergarten, and transitional first grade were also added. "Early childhood program/day care center" was added as a category for type of school. The 1993–94 PSS questionnaire items concerning types of prekindergarten programs and the number of prekindergarten teachers were deleted.

In the 1997–98 PSS, the following items were added to the survey instrument: (1) whether or not the school is coeducational (if yes, the number of male students; if no, whether the school is all female or all male); and (2) whether or not the school has a library or library media center.

There were few changes in the 1999–2000 PSS. One religious affiliation—Church of God in Christ—was added, and three associations were added—Association of Christian Teachers and Schools, National Coalition of Girls' Schools, and state or regional independent school associations. The item that previously collected data on the number of graduates that applied to 2-year or 4-year colleges was changed to collect data on the percentage of graduates who went on to attend three types of schools: 2-year colleges, 4-year colleges, and technical or other specialized schools. There also was a minor change in the definition of community type. Beginning with the 1999–2000 PSS, schools that were "rural within a Metropolitan Statistical Area" were included in the "Rural/small town" community type, while prior to the 1999–2000 PSS they were included in the "Urban fringe/large town" community type.

The 2001–02 PSS questionnaire content was relatively unchanged from the 1999–2000. One question was added to item 2 (the screener item)—"Is the school named on the front of this questionnaire located in the United States?" This question was added to facilitate the exclusion of schools from the PSS that were located outside of the United States, but had been added during the list building or area search because the school had an office with an address in the United States.

Additionally, in order to test the feasibility and benefits of collecting PSS data over the Internet, the 2001–02 PSS included an Internet response option test. The final response rate for Internet submissions was 15.4 percent for schools that received the option (5.1 percent of all schools).

Changes made to the 2003–04 PSS were minor and involved frame creation methodology, data collection procedures, and weighting procedures. For example, whereas in the 2001–02 PSS, the base weight for area frame schools was equal to the inverse of the probability of selecting the PSU in which the school resided, in the 2003–04 PSS, the base weight for area frame schools also contained a nonunitary subsampling factor for schools named solely in non-Roman Catholic religious institution lists.

Caution, however, should be used in comparing 2003–04 PSS community type estimates to those of previous years. Although the definition of community type remained unchanged, the 2003–04 PSS community types are based on the Consolidated Statistical Area/Core-Based Statistical Area rather than on the Standard Metropolitan Statistical Area/Metropolitan Statistical Area, which was used prior to the 2003–04 PSS. Also, community type is based on 2000 census data; prior to the 2003–04 PSS, community type was based on 1990 census information.

There were few changes in the 2005–06 PSS. One religious affiliation—Church of the Nazarene—was added. Also, the 2005–06 PSS used the new 12-level urban-centric locale codes, rather than the 8-level locale codes based on the Core-Based Statistical Area.

There was one change in the 2007-08 PSS. In 2005–06, non-Roman Catholic religious institutions were contacted during the area-frame operation while in 2007–08 they were not.

Future Plans

The PSS will continue as a biennial survey.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Only the area frame contributes to the standard error in the PSS. The list frame component of the standard error is always 0. Estimates of standard errors are computed using half-sample replication.

Because the area frame sample of PSUs is small (125 out of a total of approximately 2,000 eligible PSUs), there is a potential for unstable estimates of standard errors. This is particularly true when the domain of interest is small and there may not be enough information to compute a standard error. Stabilizing the standard error estimate given the level of detail of the PSS estimates would require a much larger PSU sample. The current area frame is designed to produce regional estimates.

Nonsampling Error

Coverage error. Undercoverage in the list and area frames is one possible source of nonsampling error. Because the PSS uses a dual-frame approach, it is possible to estimate the coverage, or completeness, of the PSS. A capture-recapture methodology is used to estimate the number of private schools in the United States and to estimate the coverage of private schools. In the 2003–04 PSS, the conservative coverage rate for traditional private schools was equal to 96 percent; for K-terminal private schools, it was equal to 85 percent. In the 2005–06 PSS, the overall coverage rate was 98 percent. In the 2007-08 PSS, the conservative coverage rate for traditional private schools was equal to 96 percent; for K-terminal private schools, it was equal to 93 percent.

A study comparing the quality of PSS frame coverage to that of the commercial Quality Education Data database of schools is discussed in Lee, Burke, and Rust (2000).

Nonresponse error. There are two types of nonresponse error: unit nonresponse and item nonresponse.

Unit nonresponse. In the 2007–08 PSS, the survey data from the area frame component were weighted to reflect the sampling rates (probability of selection) of the PSUs. Survey data from both the list and area frame components were adjusted for school nonresponse. There were 28,450 interviews and 2,527 cases that were noninterviews. After weighting the area frame component, these became 30,748 interviews and 2,992 noninterviews—the weighted response rate was 91 percent. In the 2005–06 PSS, the survey data from the area frame component were weighted to reflect the sampling rates (probability of selection) of the PSUs. Survey data from both the list and area frame components were adjusted for school nonresponse. There were 29,784 interviews and 1,867 cases that were noninterviews. After weighting the area frame component, these became 32,865 interviews and 2,159 noninterviews—the weighted response rate was 94 percent. In the 2003–04 PSS, of the 41,184 schools included (both traditional and K-terminal), some 9,336 cases were considered out-of-scope (that is, not eligible for the PSS). A total of 30,071 private schools completed a PSS interview, while 1,777 schools refused to participate, resulting in an overall unweighted response rate of 94 percent. When the area frame schools were weighted by the inverse of the probability of selection, the weighted response rate was 94 percent as well. In the 2001–02 PSS, the weighted response rate for traditional schools was 95 percent (96 percent unweighted); for K-terminal schools, the response rates were 97 and 96 percent, respectively. In 1999–2000, both the weighted and unweighted response rates were 93 percent for traditional schools; they were 99 and 98 percent, respectively, for K-terminal schools.

Item nonresponse. In the 2007–08 PSS, all of the weighted response rates were greater than 85 percent. The weighted item response rates for all but one variable—the percentage of graduates who went to 2-year colleges—were greater than 85 percent in 2005–06. In the 2003–04 PSS, all of the weighted response rates were greater than 85 percent. In the 2001–02 PSS, for traditional schools, all but three items had weighted response rates greater than 90 percent. The three lower rates (ranging from 77.5 percent to 86.3 percent) pertained to the percentage of graduates who went to 4-year colleges, 2-year colleges, and technical or other specialized schools. Values for items with missing data were imputed to compensate for item nonresponse.

Measurement error. NCES seeks to minimize measurement error by developing survey content in consultation with representatives of private school associations, reviewing extensively the questionnaire and instructions before distribution, requiring that the data that are not scanned are 100 percent key-verified, and processing the survey data through a comprehensive series of edits to verify accuracy and consistency.

Intersurvey Consistency in NCES Private School Surveys

The PSS and the private school component of SASS were fielded in the same school year for the first time in 1993–94. Even though these two surveys measure some of the same variables (schools, teachers, and

students), the 1993–94 results were not in agreement due to sampling and other errors. PSS results are likely to be the more accurate since the PSS serves as the sampling frame for the SASS private school component (a sample of around 3,000 schools). Special methodological studies of these two surveys have been done, including comparisons among statistical and computational procedures aimed at achieving consistency between the estimates of private schools, private school teachers, and private school students in the 1993–94 PSS and in the 1993–94 SASS—see Scheuren and Li (1995, 1996).

Data Comparability

While changes to survey design and content generally result in improved data quality, they also impact the comparability of data over time. Recent changes to the PSS and to the comparability of PSS data (both within the PSS itself and with other data sources) are discussed below.

Design change. Changes in the survey design of the 1995–96 PSS resulted in an increased number of private schools in the survey population. First, seven new association lists were obtained, adding 512 new schools to the list frame. In previous years, the area frame was relied upon to include these schools. Second, the area search results were not strictly comparable to those in previous years due to procedural differences. The 1995–96 PSS was the first survey to verify the control of schools marked as public in the screener item. Final determination of school control was based on a review of the school's name and other identifying information. As a result, several schools that had been marked as public (but which were obviously private) were added back into the PSS. They were counted as interviews if the required data were provided or as noninterviews if the required data were missing. Third, the eligibility criteria for the PSS were changed to no longer require schools to have 160 days in the school year or to conduct classes for at least 4 hours per day. Fourth, the PSS definition of a school was expanded to include programs where kindergarten is the highest grade (K-terminal schools). Additional lists of programs that might have a kindergarten were requested from nontraditional sources, and the area search was expanded to search for programs with a kindergarten. Some schools meeting the traditional PSS definition of a school (any of grades 1–12 or comparable ungraded levels) were discovered in these lists. When added to the PSS, these schools also increased the estimates of traditional schools.

Note that even when the population of schools is about the same from one survey to the next, it may represent a different set of schools. For example, the number of schools was around 27,000 in both 1997–98 and 1999–2000, although about 1,700 schools were added to the PSS universe in 1999–2000. This suggests that a nearly equal number of schools dropped out of the universe between 1997–98 and 1999–2000. Comparisons of the 1999–2000 PSS private school estimates with those from the 2001–02 PSS, however, show an overall increase in the number of private schools between 1999–2000 and 2001–02 (to about 29,000).

Questionnaire changes. Several modifications have been made to the format and content of the PSS questionnaire since 1991–92. A number of items were added (including race/ethnicity of students), and some items were deleted or modified.

Comparisons within the PSS. The estimated number of schools decreased between 2005–06 and 2007–08 (by 1,314 schools). The estimated number of private students and full-time-equivalent (FTE) teachers in 2007–08 were not statistically different from those of 2005–06. The estimated number of private schools and students decreased between the 2001–02 and 2003–04 PSS data collections (by 889 schools and 218,741 students). The estimated number of FTE teachers in 2003–04 was not statistically different from that in 2001–02. Comparisons of the 2001–02 PSS estimates with those from previous PSS data collections show increases in the number of private schools, students, and teachers between 1999–2000 and 2001–02. Comparisons of the 1999–2000 PSS estimates with those from previous surveys show no significant change in the estimated number of private schools; however, they do indicate an increase in the estimated number of private school teachers and students.

Comparisons with the Current Population Survey. A comparison of the PSS estimate of K–12 students enrolled in all private schools (traditional and k-terminal) with the household survey estimate from the 2007 October Supplement of the Current Population Survey (CPS) (U.S. Census Bureau 2006) shows that the PSS estimate of 5,072,451 does statistically differ from the CPS estimate of the number of private school students in grades kindergarten through 12 in October 2007 of 4,817,000. A comparison of the 2003–04 PSS estimate of K–12 students enrolled in all private schools (traditional and K-terminal) with the household survey estimate from the 2003 October Supplement to the CPS shows that the PSS estimate of 5,212,992 students is not statistically different from the CPS estimate of 5,259,000 students (U.S. Census Bureau 2005). A comparison of the 2001–02 PSS estimate of K–12 students enrolled in all private schools (traditional and K-terminal) with the household survey estimate from the October 2001 CPS shows that the PSS estimate of 5,439,925 is higher than the CPS estimate of 5,164,000; the 95 percent confidence interval of the PSS estimate ranges from 5,383,898 to 5,495,952 students, while that of the CPS estimate ranges from 4,956,000 to 5,372,000 students. In the 1995–96 school year, the PSS and CPS estimates did

not differ significantly; in 1997–98, the PSS estimate was higher than the CPS estimate; and, in 1999–2000, the PSS estimate was lower than the CPS estimate. Comparisons between CPS and PSS enrollment estimates for earlier years are not as informative since, prior to 1995–96, the PSS estimates did not include the kindergarten enrollment from K-terminal schools, whereas the CPS has always included it.

Comparisons with National Catholic Educational Association data. Comparisons of the PSS estimates for Catholic schools, students, and FTE teachers (traditional schools) with the National Catholic Educational Association (NCEA) (National Catholic Educational Association 2008) data for the 2007–08 school year show differences in the school (7,507 versus 7,378), student (2,156,173 versus 2,270,913) and FTE teacher counts (146,627 versus 160,075) between PSS and NCEA, respectively. Comparisons of the PSS estimates for Catholic schools, students, and FTE teachers with the NCEA data for the 2003–04 school year show differences in the number of students (2,365,220 vs. 2,484,252) and FTE teachers (152,611 vs. 162,337) between PSS and NCEA, respectively. The difference between the PSS estimate of 7,919 Catholic schools and the NCEA count of 7,955 schools is not statistically significant. The survey methodologies used by NCES and NCEA are quite different; NCES surveys private schools directly, while NCEA surveys archdiocesan and diocesan offices of education and some state Catholic conferences. The NCEA and PSS computations of full-time equivalents differ in the weight assigned to part-time teachers; thus, the PSS and NCEA counts of FTE teachers are not strictly comparable.

For the 2001–02 school year, comparisons of the PSS estimate for Catholic schools with the NCEA data show differences in the school and student counts. The NCEA count of 8,000 schools is below the lower limit of the 95 percent confidence interval of the PSS estimate of Catholic schools (which ranges from 8,112 to 8,302). The NCEA K–12 student count of 2,553,277 is higher than the upper limit of the 95 percent confidence interval of the PSS estimate of Catholic students (which ranges from 2,492,773 to 2,538,274). Both the NCEA teacher count of 163,004 and the PSS estimate of 155,514 include part- and full-time teachers in the computation of full-time equivalents (the 95 percent confidence interval of the PSS estimate ranges from 153,902 to 157,126).

NCES publication criteria for the PSS. NCES criteria for the publication of an estimate are dependent on the type of survey—sample or universe. To publish an estimate for a sample survey, at least 30 cases must be used in developing the estimate. For a universe survey, a minimum of three cases must be used. The PSS includes both types of surveys: (1) a sample survey of PSUs (area frame) that collects data on schools not in the list frame (the number of PSUs changes for each administration); and (2) a complete census of schools belonging to the list frame. NCES has established a rule that published PSS estimates must be based on at least 15 schools. If the estimate satisfies this criterion and the coefficient of variation (standard error/estimate) is greater than 25 percent, the estimate is identified as having a large coefficient of variation and the reader is referred to a table of standard errors.

6. CONTACT INFORMATION

For content information on the PSS, contact:

Stephen Broughman
Phone: (202) 502-7315
E-mail: stephen.broughman@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Broughman, S.P. (1996). *Private School Universe Survey, 1993–94* (NCES 96-143). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Broughman, S.P., and Colaciello, L.A. (1998). *Private School Universe Survey, 1995–96* (NCES 98-229). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Broughman, S.P., and Colaciello, L.A. (1999). *Private School Universe Survey, 1997–98* (NCES 1999-319). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Broughman, S.P., and Colaciello, L.A. (2001). *Private School Universe Survey, 1999–2000* (NCES 2001-330). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Broughman, S.P., Gerald, E., Bynum, L.T., and Stoner, K. (1994). *Private School Universe Survey, 1991–92* (NCES 94-350). National Center

for Education Statistics, U.S. Department of Education. Washington, DC.

Broughman, S.P., and Pugh, K.W. (2004). *Characteristics of Private Schools in the United States: Results From the 2001–2002 Private School Universe Survey* (NCES 2005-305). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Broughman, S.P., and Swaim, N.L. (2006). *Characteristics of Private Schools in the United States: Results From the 2003–2004 Private School Universe Survey* (NCES 2006-319). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Broughman, S.P., Swaim, N.L., and Keaton, P.W. (2008). *Characteristics of Private Schools in the United States: Results From the 2005–2006 Private School Universe Survey* (NCES 2008-315). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Broughman, S.P., Swaim, N.L., and Keaton, P.W. (2009). *Characteristics of Private Schools in the United States: Results From the 2007-08 Private School Universe Survey* (NCES 2009-313). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Gerald, E., McMillen, M., and Kaufman, S. (1992). *Private School Universe Survey, 1989–90* (NCES 93-122). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

National Catholic Educational Association. (2008). *United States Catholic Elementary and Secondary Schools, 2007–08: Annual Statistical Report on Schools, Enrollment, and Staffing*. Washington, DC: Author.

U.S. Census Bureau. (2005). School Enrollment-Social and Economic Characteristics of Students: October 2003. Retrieved March 24, 2005, from http://www.census.gov/population/www/socdemo/school/spw2003.html.

U.S. Census Bureau. (2006). School Enrollment-Social and Economic Characteristics of Students: October 2005. Retrieved March 24, 2009, from http://www.census.gov/population/www/socdemo/school/cps2005.html.

Survey Design

McMillen, M., and Benson, P. (1992). *Diversity of Private Schools* (NCES 92-082). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Scheuren, F., and Li, B. (1995). *Intersurvey Consistency in NCES Private School Surveys* (NCES Working Paper 95-16). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Scheuren, F., and Li, B. (1996). *Intersurvey Consistency in NCES Private School Surveys for 1993–94* (NCES Working Paper 96-27). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Broughman, S., Tourkin, S., Swaim, N., Hryczaniuk, C., Guan, X., Peterson, J., Parmer, R., Abramson, R., Zhang, B., Andriani, S., & Walter, E. (2009). *2007–08 Private School Universe Survey (PSS) Data File User's Manual and Survey Documentation* (NCES 2009-319). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Jackson, B.J., and Frazier, R.L. (1996). *Improving the Coverage of Private Elementary-Secondary Schools* (NCES Working Paper 96-26). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jackson, B.J., Johnson, N.R., and Frazier, R.L. (1997). Improving the Coverage of Private Elementary-Secondary Schools. In *Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association* (NCES Working Paper 97-41). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Lee, H., Burke, J., and Rust K. (2000). Evaluating the Coverage of the U.S. National Center for Education Statistics' Public and Private School Frames Using Data From the National Assessment of Educational Progress. In *Proceedings of the Second International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions* (pp. 89–98). Arlington, VA: American Statistical Association.

Tourkin, S., Swaim, N., Guan, X., Peterson, J., Schwanz, D., Warner, T., Parmer, R., Wilson, M., Broughman, S., and Walter, E. (2008). *2003-04 Private School Universe Survey (PSS) Data File User's Manual and Survey Documentation* (NCES

2008-314). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 4: Schools and Staffing Survey (SASS)

1. OVERVIEW

The NCES Schools and Staffing Survey (SASS) provides data on public and private schools, principals, school districts, and teachers. SASS gathers information about many topics, including various characteristics of elementary and secondary students, some of the professional and paraprofessional staff who serve them, the programs offered by schools, principals' and teachers' perceptions of school climate and problems in their schools, teacher compensation, and district hiring practices. SASS is a unified set of surveys that facilitates comparisons between public and private schools and allows linkages of teacher, school, school district, and principal data. First conducted in school year 1987–88, SASS has been conducted six times, most recently in school year 2007–08.

SAMPLE SURVEY OF PUBLIC, PRIVATE, CHARTER, AND BIE SCHOOLS

SASS collects data on:

- School districts
- Principals
- Schools
- Teachers
- Library media centers

Purpose

The purpose of SASS is to collect the information necessary for a complete picture of American elementary and secondary education. SASS is designed to provide national estimates of public elementary, secondary, and combined schools and teachers; state estimates of public elementary and secondary schools and teachers; and estimates for private schools; teachers and principals at the national level; and by private school affiliation. The SASS questionnaires were revised for the 2003–04 and the 2007–08 administrations, with the addition of new items about teachers' career paths, parental involvement, school safety, and institutional support for information literacy. The questionnaires continued to measure the same five policy issues: teacher shortage and demand; characteristics of elementary and secondary teachers; teacher workplace conditions; characteristics of school principals; and school programs and policies.

Core Components

SASS consists of four core components administered to districts, schools, principals, and teachers. The district questionnaire is sent to a sample of public school districts. The school questionnaire is sent to a sample of public schools and private schools, as well as all charter schools in operation as of 1998–99, and all schools operated by the Bureau of Indian Education (BIE) or American Indian/Alaska Native tribes. The principal and teacher questionnaires are sent to a sample of principals and teachers working at the schools that receive the school questionnaire. (The Teacher Follow-up Survey is a fifth component of SASS and is covered in chapter 5.)

School District Survey (formerly the Teacher Demand and Shortage Survey). The questionnaire for this survey is mailed to each sampled local education agency (LEA). The respondents are contact people identified by LEA personnel.

If no contact person is identified, the questionnaire is addressed to “Research Director.” The School District Questionnaire consists of items about student enrollment, number of teachers, teacher recruitment and hiring practices, teacher dismissals, existence of a teacher union, length of the contract year, teacher compensation, school choice, magnet programs, graduation requirements, oversight of home-schooled students and charter schools, use of school performance reports, migrant education, and professional development for teachers and administrators. Some items that appeared previously have been dropped, such as those that collected layoff data and counts of students by grade level (the latter are available through the NCES Common Core of Data [CCD]). In the 2003–04 administration, new topics, including principal hiring practices and instructional aide hiring practices, were added to the questionnaire. In the 2007–08 administration, items on district performance, teacher tenure and dismissal, principal salary, length of the contract year for teachers, and type of retirement benefits for teachers were added or revised.

The School District Questionnaire is mailed only to public school districts. Independent public charter schools, BIE-funded schools, and schools that are the only school in the district are given the School Questionnaire (with district items), not the School District Questionnaire. The School Questionnaire (with district items) includes all of the items included in the School Questionnaire as well as selected items from the School District Questionnaire. The applicable items for private schools appear in the Private School Questionnaire.

School Principal Survey (formerly the School Administrator Survey). The questionnaire for this survey collects information about principal/school head demographic characteristics, training, experience, salary, and judgments about the seriousness of school problems. Information is also obtained on professional development opportunities for teachers and principals, teacher performance, barriers to dismissal of underperforming teachers, school climate and safety, parent/guardian participation in school events, and attitudes about educational goals and school governance. The 2007–08 questionnaire appeared in two versions: one for principals or heads of public schools and one for heads of private schools. The two versions contain minor variations in phrasing to reflect differences between public and private schools in governing bodies and position titles in schools. Items on experience prior to becoming a principal, teacher and school performance, and time allocation for students during the week were added or revised in the 2007–08 questionnaire.

School Survey. The questionnaires for this survey are sent to public schools, private schools, BIE schools, and charter schools. Private schools receive the Private School Questionnaire, while BIE schools and charter schools receive the School Questionnaire (with district items), described separately below. As in 2003–04, the 2007–08 data collection for the private school component of SASS coincided with the administration of the NCES Private School Universe Survey (PSS). Since both PSS and SASS were administered in 2007–08, to reduce respondent burden, the private schools in the SASS sample were not sent a PSS questionnaire. Instead, the PSS items appeared in the SASS Private School Questionnaire. (See chapter 3 for a complete description of PSS.)

The School Questionnaire is addressed to “Principal,” although the respondent can be any knowledgeable school staff member (e.g., vice principal, head teacher, or school secretary). Items cover grades offered, student attendance and enrollment, staffing patterns, teaching vacancies, high school graduation rates, programs and services offered, curriculum, and college application rates. The Private School Questionnaire also includes items from the School District Questionnaire that are applicable to private schools. The 2007–08 collection included items on the beginning time of students’ school day; length of the school year for students; school websites; and math, reading, or science specialist assignments.

School Questionnaire (with district items). The purpose of the questionnaire (which was also referred to as the Unified School Questionnaire in the 2003–04 SASS) was to obtain information about schools, such as grades offered, number of students enrolled, staffing patterns, teaching vacancies, high school graduation rates, programs and services offered, and college application rates. Schools that are the only school in the district, state-run schools (e.g., schools for the blind), charter schools that do not report to a traditional school district, and BIE-funded schools received the School Questionnaire (with district items), an expanded version of the Public School Questionnaire that included items from the School District Questionnaire.

Teacher Survey. The questionnaire for this survey is mailed to a sample of teachers from the SASS sample of schools. It is sent to teachers in public schools, private schools, charter schools, and BIE schools. The Teacher Questionnaire collects data from teachers about their education and training, teaching assignment, certification, workload, and perceptions

and attitudes about teaching. Questions are also asked about teacher preparation, induction, organization of classes, computers, and professional development. The only eligible respondent for each teacher questionnaire is the teacher named on the questionnaire label. As of the 1993–94 SASS, administrators are eligible for both the Teacher Survey and the Principal Survey, if they teach a regularly scheduled class. In the 2007–08 Teacher Survey, items on grade range of teaching certification, use of electronic communications with parents, and out-of-pocket expenses for school supplies were added or revised.

Teacher Listing Form. The SASS Teacher Listing Form collects the full list of teachers from a school, along with information on subject matter taught, full- or part-time teaching status, and teaching experience. A question about teachers' race/ethnicity was replaced in the 2007–08 data collection by a question about teachers' status for the next school year. The information in the Teacher Listing Form is used to select a representative teacher sample and send out the Teacher Questionnaires. In 2007–08, the Teacher Listing Form restored a section that was removed in 2003–04, which had asked about the school name and grade range for verification purposes. (This section was not included in the survey questionnaire in 2003–04, as it was verified at the school, using a laptop-collected form.)

Additional Components

In addition to the core data collection described above, SASS featured additional components focusing on library media specialists/librarians and on student records in 1993–94 and on library media centers in 1993–94, 1999–2000, 2003–04, and 2007–08. One year following each SASS, a Teacher Follow-up Survey (TFS) is mailed to a sample of participants in the SASS Teacher Survey. (See chapter 5 for a complete description of TFS.) In 2007-08, SASS also included a Principal Follow-up Survey.

School Library Media Center Survey. This survey was added in the 1993–94 SASS. The questionnaire for the survey asks public and BIE schools about their access to and use of new information technologies. The questionnaire was not sent to private schools in 2003–04, due to budgetary reasons. (In 2007–08, the survey only surveyed public schools as well.) The survey collects data on library collections, media equipment, use of technology, staffing, student services, expenditures, currency of the library collection, and collaboration between the library media specialist and classroom teachers. A section on information literacy was added to the 2003–04 questionnaire. Items on access to online licensed databases, resource availability, and information literacy were added or revised in the 2007–08 questionnaire. (See chapter 10 for a more complete description of this survey.)

School Library Media Specialist/Librarian Survey. The questionnaire for this survey was mailed to a subsample of the SASS sample of public, private, and BIE schools in 1993–94. The survey solicited data that could be used to describe school librarians—for example, their educational background, work experience, and demographic characteristics. Because much of the collected information was comparable to that obtained in the Teacher Questionnaire, comparisons between librarians and classroom teachers can be made.

Periodicity

Between the 1987–88 and 1993–94 school years, SASS core components were on a 3-year cycle, with the TFS conducted 1 year after SASS. After a 6-year hiatus, SASS was fielded again in the 1999–2000, 2003–04, and 2007–08 school years (with the TFS following in 2000–01, 2004–05, and 2009–10). Subsequent SASS administrations are scheduled on a 4-year cycle.

2. USES OF DATA

SASS is the largest, most extensive survey of school districts, schools, principals, teachers, and library media centers in the United States today. It includes data from the public, private, and BIE school sectors. Moreover, SASS is the only survey that studies the complete universe of public charter schools. Therefore, SASS provides a multitude of opportunities for analysis and reporting on issues related to elementary and secondary schools.

SASS data have been collected six times between 1987 and 2007. Many questions have been asked of respondents at multiple time points, allowing researchers to examine trends on these topics over time. SASS asks similar questions of respondents across sectors, including public, public charter, BIE, and private schools. The consistency of questions across sectors and the large sample sizes allow for exploration of similarities and differences across sectors.

SASS data are representative at the state level for public school respondents and at the private school affiliation level for private school respondents. Thus, SASS is invaluable for analysts interested in elementary, middle, and secondary schools within or across specific states or private school affiliations. The

large SASS sample sizes allow extensive disaggregation of data according to the characteristics of teachers, administrators, schools, and school districts. For example, researchers can compare urban and rural settings and the working conditions of teachers and administrators of differing demographic backgrounds.

SASS collects extensive data on teachers, principals, schools, and school districts. Information on teachers includes their qualifications, early teaching experience, teaching assignments, professional development, and attitudes about the school. The SASS School Principal Questionnaire collects information about principals' or school heads' years of experience and training, goals and decision making, professional development for teachers and instructional aides, school climate and safety, student instructional time, principal perceptions and working conditions, and demographic information. Questions about schools include enrollment, staffing, the types of programs and services offered, school leadership, parental involvement, and school climate. At the district level, information is sought on the recruitment and hiring of teachers, professional development programs, student services, and other relevant topics.

SASS data can be very useful for researchers performing their own focused studies on smaller populations of teachers, administrators, schools, or school districts. SASS can supply data at the state, affiliation, or national level that provide valuable contextual information for localized studies; localized studies can provide illustrations of broad findings produced by SASS.

Users of restricted-use SASS data can link school districts and schools to other data sources. For instance, 2007–08 SASS restricted-use datasets include selected information taken from the CCD, but researchers can augment the datasets by adding more data from the CCD—either fiscal or nonfiscal data.

3. KEY CONCEPTS

Because of the large number of concepts in SASS surveys, only those pertaining to the level of data collection (LEA, school, teacher, library) are described in this section. For additional terms, the reader is referred to glossaries in SASS reports.

Local Education Agency (LEA). A public school district, or LEA, is defined as a government agency employing elementary- and secondary-level teachers and administratively responsible for providing public elementary and/or secondary instruction and educational support services. Districts that do not operate schools but employ teachers were last included in the 1999–2000 SASS. (For example, some states have special education cooperatives that employ special education teachers who teach in schools in more than one school district.)

Public School. An institution that provides educational services for at least one of grades 1–12 (or comparable ungraded levels), has one or more teachers to give instruction, is located in one or more buildings, receives public funds as primary support, and is operated by an education agency. Schools in juvenile detention centers and schools located on military bases and operated by the Department of Defense are included.

Private School. An institution that is not in the public system and that provides instruction for any of grades 1–12 (or comparable ungraded levels). The instruction must be given in a building that is not used primarily as a private home. Private schools are divided into three categories: (1) Catholic: parochial, diocesan, private order; (2) other religious: affiliated with a conservative Christian school association, affiliated with a national denomination, unaffiliated; and (3) nonsectarian: regular, special program emphasis, special education. The classification of nonsectarian schools by program emphasis disentangles private schools offering a conventional academic program (regular) from those that either serve special-needs children (special education) or provide a program with a special emphasis (e.g., arts and sciences).

Charter School. A charter school is a public school that, in accordance with an enabling state statute, has been granted a charter exempting it from selected state or local rules and regulations. A charter school may be a newly created school or it may previously have been a public or private school.

BIE School. A school funded by the Bureau of Indian Education of the Bureau of Indian Affairs, U.S. Department of the Interior. These schools may be operated by the BIE, a tribe, a private contractor, or an LEA.

Library Media Center. A library media center is an organized collection of printed, audiovisual, or computer resources that (a) is administered as a unit, (b) is located in a designated place or places, and (c) makes resources and services available to students, teachers, and administrators.

Teacher. A full- or part-time teacher who teaches any regularly scheduled classes in any of grades K–12.[1] This includes administrators, librarians, and other professional or support staff who teach regularly scheduled classes on a part-time basis. Itinerant teachers are also included, as well as long-term substitutes who are filling the role of a regular teacher on a long-term basis. An itinerant teacher is one who teaches at more than one school (e.g., a music teacher who teaches 3 days per week at one school and 2 days per week at another). Short-term substitute teachers and student teachers are not included.

4. SURVEY DESIGN

Target Population

LEAs that employ elementary- and/or secondary-level teachers (e.g., public school districts, state agencies that operate schools for special student populations, such as inmates of juvenile correctional facilities or students in Department of Defense schools); cooperative agencies that provide special services to more than one school district; public, private, BIE, and charter schools with students in any of grades 1–12; the principals of these schools; library media centers; and teachers in public, private, BIE, and charter schools who teach students in grades K–12 in a school with at least a 1st grade.

Sample Design

SASS uses a stratified probability sample design. Details of stratification variables, sample selection, and frame sources are provided below.

Public school sample. In the public school sample, schools are selected first. The first level of stratification is by type of school: (a) BIE schools (all BIE schools are automatically in the sample); (b) schools with a high percentage of American Indian students (i.e., schools with 19.5 percent or more American Indian students); (c) schools in Delaware, Florida, Maryland, Nevada, and West Virginia (where it is necessary to implement a different sampling methodology to select at least one school from each LEA in the state); (d) charter schools; and (e) all other schools. Schools falling into more than one group are assigned to types A, B, D, C, and E in that order. The second level of stratification varies within school type. All BIE schools are automatically selected for the sample, so no stratification is needed. Schools with a high percentage of American Indian students are stratified by state (Arizona; California; Montana; New Mexico; Washington; the remaining western states; Minnesota; North Dakota; South Dakota; the remaining midwestern states; North Carolina; Oklahoma; and the remaining states except Alaska, since most Alaskan schools have a high Native American enrollment). Schools in Delaware, Florida, Maryland, Nevada, and West Virginia are stratified first by state and then by LEA. Charter schools and schools not placed in another category are stratified by state. Within each second level, there are three grade level strata (elementary, secondary, and combined schools).

Within each stratum, all non-BIE schools are systematically selected using a probability proportionate to size algorithm. The measure of size used for schools in the CCD is the square root of the number of teachers in the school as reported in the CCD file. Any school with a measure of size larger than the sampling interval is excluded from the probability sampling operation and included in the sample with certainty.

The CCD Public Elementary/Secondary School Universe Survey serves as the public school sampling frame. (See chapter 2 for a complete description of the CCD.) The frame includes regular public schools, Department of Defense-operated military base schools, and special purpose schools (such as special education, vocational, and alternative schools). Schools outside the United States and schools that teach only prekindergarten, kindergarten, or postsecondary students are deleted from the file. The following years of the CCD were used as the public school frame for the last five rounds of SASS:

- 2005–06 CCD for the 2007–08 SASS;
- 2001–02 CCD for the 2003–04 SASS;
- 1997–98 CCD for the 1999–2000 SASS;
- 1991–92 CCD for the 1993–94 SASS; and
- 1988–89 CCD for the 1990–91 SASS.

In the 1987–88 SASS, the 1986 Quality Education Data (QED) survey was used as the sampling frame (Kaufman 1991).

Private school sample. For private schools, the sample is stratified within each of the two types of frames: (1) a list frame, which is the primary private school frame; and (2) an area frame, which is used to identify schools not included in the list frame and to compensate for the undercoverage of the list frame. Private schools in the

[1] A teacher teaching only kindergarten students is in scope, provided the school serves students in a grade higher than kindergarten.

list frame are stratified by affiliation, grade level, and region. Within each stratum, schools are sampled systematically using a probability proportionate to size algorithm. Any school with a measure of size larger than the sampling interval is excluded from the probability sampling process and included in the sample with certainty. All schools in the area frame within noncertainty PSUs and not already listed in the list frame are included in the sample with certainty.

The most recent PSS, updated with the most recent association lists, serves as the private school sampling frame. For example, the 2001–02 PSS—updated with 26 lists of private schools provided by a private school association (as well as 51 lists of private schools, from the 50 states and the District of Columbia)—was used as the private school frame for the 2003–04 SASS. For the 2007–08 SASS, the private school list frame was based on the 2005–06 PSS, updated with private school organizations and state lists collected by the U.S. Census Bureau in the summer of 2006. The 1991–92, 1989–90, and 1997–98 PSS were the basis for the private school frame for the 1993–94, 1990–91, and 1999–2000 SASS, respectively. The 1986 QED survey was used as the sampling frame for the 1987–88 SASS.

BIE school selection. Since the 1993–94 SASS, all BIE schools have been selected with certainty; in 1990–91, 80 percent of BIE schools were sampled. The BIE school frame for the 2003–04 SASS consisted of a list of schools that the BIE operated or funded during the 2001–02 school year. (The list was obtained from the U.S. Department of the Interior.) The BIE list was matched against the CCD, and the schools on the BIE list that did not match the CCD were added to the universe of schools.

For the 2007–08 SASS data collection, a separate universe of schools operated or funded by the BIE in the 2005–06 school year was drawn from the Program Education Directory maintained by the BIE. (The CCD now defines the BIE as its own "territory," similar to Puerto Rico and other non-state territories, and does not permit duplicates to be reported by the states.) All BIE schools meeting the SASS definition of a school were included in the sample.

Charter school selection. In the 1999–2000 SASS, a charter school sample was added. All charter schools were selected with certainty from the frame, which consisted of a list of charter schools developed for the U.S. Department of Education's Institute of Education Sciences. The list included only charter schools that were open (teaching students) during the 1998–99 year. This changed in the 2003–04 SASS, when a nationally representative sample of public charter schools was included as part of the public school sample. In the 2007–08 SASS, charter schools continued to be included as a part of the public school sample.

Each school sampled for SASS receives a school questionnaire, and the principal of each sampled school receives a principal questionnaire.

Teacher selection. Within each sampled school, a sample of teachers is selected. First, the sampled schools are asked to provide a list of their teachers and selected characteristics. For example, in the 2007–08 SASS data collection, the Teacher Listing Form was collected as early as possible in the 2007–08 school year at all public (including public charter), private, and BIE-funded schools in the SASS sample to obtain a complete list of all the teachers employed at each school.

In the 2007–08 SASS, teachers were stratified into one of two teacher types: new and experienced. For new and experienced teachers in public schools, oversampling was not required, due to the large number of sampled schools with new teachers. Therefore, teachers were allocated to the new and experienced categories in proportion to their numbers in the school. However, in private schools, new teachers were oversampled. Before teachers were allocated to the new or experienced strata, schools were first allocated an overall number of teachers to be selected.

Teacher records within a school are sorted by the teacher stratum code, the teacher subject code, and the teacher line number code. The teacher line number code is a unique number assigned to identify the teacher within the list of teachers keyed by the field representative. Within each teacher stratum in each school, teachers are selected systematically with equal probability. The within-school probabilities of selection are computed so as to give all teachers within a school stratum the same overall probability of selection (self-weighted) within teacher and school strata, but not across strata. However, since the school sample size of teachers is altered due to the minimum constraint (i.e., at least one teacher per school) or maximum constraint (i.e., no more than either twice the average stratum allocation or 20 teachers per school), the goal of achieving self-weighting for teachers is lost in some schools. Each sampled teacher receives a teacher questionnaire.

Library media center selection. For the 2003–04 and 2007–08 SASS, all library media centers in public, public charter, and BIE-funded schools in the SASS

sample were asked to complete the School Library Media Center Questionnaire.

School district selection. In most states, once public schools are selected, the districts associated with these schools are placed in the sample as well. However, in Delaware, Florida, Maryland, Nevada, and West Virginia, all districts are defined as school sampling strata, placing all districts in each of these states in the district sample. (In some SASS administrations, a sample of districts not associated with schools is taken, but not in the 2007–08 SASS.) The district sample is selected using a probability proportionate to size algorithm. Each sampled school district receives a school district questionnaire.

The approximate sample sizes for the 2007–08 SASS were 12,900 schools and administrators, some 56,370 teachers, and 5,250 school districts.

Data Collection and Processing

The 2007–08 SASS was primarily a mailout/mailback survey with computer-assisted telephone interviewing (CATI) and telephone follow-up. In 2003–04 and 2007–08, the School Library Media Center Survey did not have an Internet reporting option, as it did in 1999–2000. All survey modes used in SASS are administered by the U.S. Bureau of the Census.

Reference Dates. Data for SASS components are collected during a single school year. Most data items refer to that school year. Questions on enrollment and staffing refer to October 1 of the school year. Questions for teachers about current teaching loads refer to the most recent full week that school was in session, and questions on professional development refer to the past 12 months.

Data Collection. The data collection procedures begin with advance mailings to school districts explaining the nature and purpose of SASS. Field staff then phone school principals to set up face-to-face appointments with them. The telephone call includes a request to prepare a list of all eligible teachers in their schools. If the teacher roster is not provided at the appointment, field staff make arrangements to obtain the roster at a later meeting. The teacher sample is selected using these lists.

The school district questionnaires are mailed out first. Then, the school, principal, and library media center surveys are delivered to schools in person. The teacher questionnaires are delivered last. Follow-up efforts begin approximately 2 weeks after questionnaires are distributed. They consist of telephone calls and personal visits to schools to obtain completed questionnaires or to verify that they have been mailed back. Field staff record the status of each questionnaire and, if necessary, supply additional blank questionnaires.

Processing. During the check-in phase, each questionnaire is assigned an outcome code: completed interview, out-of-scope, or noninterview. A combination of manual data keying and imaging technology was used to enter the data. Then, interview records in the data files undergo a round of primary data review, where analysts examine the frequencies of each data item in order to identify any suspicious values. Census staff review the problem cases and make corrections whenever possible.

After the primary data review, all records (i.e., records from all survey components) classified as interviews are subject to a set of computer edits: a range check, a consistency edit, and a blanking edit. After the completion of these edits, the records are put through another edit to make a final determination of whether the case is eligible for the survey, and, if so, whether sufficient data have been collected for the case to be classified as an interview. A final interview status recode (ISR) value is assigned to each case as a result of the edit.

Estimation Methods

Sample units are weighted to produce national and state estimates for public elementary and secondary school surveys (i.e., schools, teachers, administrators, school districts, and school library media centers); and national estimates for BIE, charter school, and public combined school surveys (i.e., schools, teachers, administrators, and school library media centers). The private sector is weighted to produce national and affiliation group estimates. These estimates are produced through the weighting and imputation procedures discussed below.

Weighting. Estimates from SASS sample data are produced by using weights. The weighting process for each component of SASS includes adjustments for nonresponse using respondents' data and adjustments of the sample totals to the frame totals to reduce sampling variability. The exact formula representing the construction of the weight for each component of SASS is provided in each administration's sample design report (e.g., *1993–94 Schools and Staffing Survey: Sample Design and Estimation* [Abramson et al. 1996]). The construction of weights is also discussed in the *Quality Profile* reports (Jabine 1994; Kalton et al. 2000) and in the documentation for the 2003–04 administration (Tourkin et al. 2007). Since SASS and PSS data were collected at the same time in

1993–94 and 1999–2000, in both years the number of private schools reported in SASS was made to match the number of private schools reported in PSS.

Imputation. In all administrations of SASS, all items with missing values are imputed for records classified as interviews. SASS uses a two-stage imputation procedure. The first-stage imputation uses a logical or deductive method, such as:

- Using data from other items in the same questionnaire;
- Extracting data from a related SASS component (different questionnaire); or
- Extracting information about the sample case from the PSS or CCD, the sampling frames for private and public schools, respectively.

In addition, some inconsistencies between items are corrected by ratio adjustment during the first-stage imputation.

The second-stage imputation process is applied to all items with missing values that were not imputed in the first stage. This imputation uses a hot-deck imputation method, extracting data from a respondent (i.e., a donor) with similar characteristics to the nonrespondent. If there is still no observed value after collapsing to a certain point, the missing values are imputed using a clerically imputed value or automated algorithm.

Recent Changes

Several changes were made over time, largely due to budgetary reasons.

Design changes from 1999–2000 to 2007–08:

- Rather than surveying all public charter schools, as was done in the 1999–2000 SASS, some 300 public charter schools were sampled for the 2003–04 SASS.
- The separate questionnaire for public charter schools was discontinued. The reduction in the public charter school sample size from 1,100 in the 1999–2000 SASS to about 300 in the 2003–04 SASS meant it was no longer feasible to produce a separate questionnaire, since public charter school data could not be published with as much detail (for the 2003–04 SASS, only at the national and regional levels). Public charter school data are now included with traditional public school data.
- Affiliation for private schools was redefined and stratified into 17 groups rather than the previous 20 groups in the 2003–04 SASS. Catholic schools were split into three groups based on typology. Other religious schools were divided into five groups corresponding to the four largest non-Catholic religious organizations (by number of schools) and a catch-all “other.” Nonsectarian schools were divided into three groups by typology.
- Grade-level stratification in public and private schools was defined purely on the basis of grade level of the school starting in 2003–04 SASS. Schools classified as a type other than “regular school” were no longer placed by default in the combined school category, which includes schools with some elementary and some secondary grades. Many nonregular schools (i.e., special education, alternative, and vocational schools) cover a specific grade range. To the extent this grade range is known, this was a more appropriate method of stratification than placing them all in the combined school strata. Nonregular schools with a grade range that is ungraded or unknown remain in the combined school strata.
- Public schools from the CCD were collapsed into what was perceived to be a better fit with the SASS definition of a school prior to stratification beginning in the 2003–04 SASS. The sample allocation was revised to avoid undersampling schools now classified at the combined grade level. In other words, the revision of the sample allocation ensured that the newly combined schools were sampled at the same approximate rate as they would have been prior to the collapsing procedure. In general, the combined school sample size was increased to the point at which the combined school sampling rate equaled the overall state-level sampling rate. For example, if one in five schools were sampled in a particular state, then one in five of the combined schools were sampled rather than using the default sample size of 10 combined schools.
- The sort order for the public and private school sampling was altered to sort on enrollment in a serpentine fashion (instead of always sorting in descending order) in the 2003–04 SASS. Serpentine sorting involves sorting in ascending order with respect to higher level sort variables one time, then sorting in descending order the next time, and so on. This

reduces the variation in enrollment between adjacent sampled schools and thus reduces the overall sampling error.

- Florida and Maryland were added to the list of states where at least one school is selected in each school district. This was done in the 2003–04 SASS to decrease the standard error of the state-level school district estimates.

- Oversampling of bilingual/English as a Second Language (ESL) teachers was discontinued in the 2003–04 SASS, since a sufficient number of bilingual teachers to produce the desired reliability estimates could be done without oversampling.

- Teacher sampling was automated to speed up the distribution of the teacher questionnaires. This, however, reduced the level of control over the sample sizes for the remaining oversampled teacher strata (Asian/Pacific Islander and American Indian/Alaska Native). The automation no longer allowed the sampling rate for these teachers to be periodically revised during the sampling process. Thus, if the number of these teachers listed differed from the expected number, the sample size goal would no longer be met.

- The School Library Media Center Questionnaire was not administered to private schools for budget reasons as of the 2003–04 SASS.

- The School Questionnaire (with district items) is a questionnaire that contains the public school questions and most of the school district questions in the 1999–2000 SASS. It was administered to public charter, state-operated (often schools for the blind or schools located in juvenile detention facilities), and BIE-funded schools, as well as public schools in one-school districts. This change was made to ease respondent burden in cases where the respondent for the school and school district questionnaires was expected to be the same.

Future Plans
SASS administrations are now scheduled on a 4-year cycle. The next administration will be in 2011–12.

5. DATA QUALITY AND COMPARABILITY

Sampling Error
The estimators of sampling variances for SASS statistics take the SASS complex sample design into account. For an overview of the calculation of sampling errors, see the *Quality Profile* reports (Jabine 1994; Kalton et al. 2000).

Direct Variance Estimators. The balanced half-sample replication (BHR) method, also called balanced repeated replication (BRR), was used to estimate the sampling errors associated with estimates from the 1987–88 and 1990–91 SASS. Given the replicate weights, the statistic of interest (e.g., the number of 12th grade teachers from the School Survey) can be estimated from the full sample and from each replicate. The mean square error of the replicate estimates around the full sample estimate provides an estimate of the variance of the statistic.

A bootstrap variance estimator was used for the 1993–94, 1999–2000, 2003–04, and 2007–08 SASS. The bootstrap variance reflects the increase in precision due to large sampling rates because the bootstrap is done systematically without replacement, as was the original sampling. Bootstrap samples can be selected from the bootstrap frame, replicate weights computed, and variances estimated with standard BHR software. The bootstrap replicate basic weights (inverse of the probability of selection) were subsequently reweighted. More information on the bootstrap variance methodology and how it applies to SASS is contained in the following sources: “A Bootstrap Variance Estimator for Systematic PPS Sampling” (U.S. Department of Education 2000) which describes the methodology used in the 1999–2000 SASS; “A Bootstrap Variance Estimator for the Schools and Staffing Survey” (U.S. Department of Education 1994); “Balanced Half-Sample Replication With Aggregation Units” (U.S. Department of Education 1994); “Comparing Three Bootstrap Methods for Survey Data” (Sitter 1990); “Properties of the Schools and Staffing Survey Bootstrap Variance Estimator” (U.S. Department of Education 1996); and “The Jackknife, the Bootstrap and Other Resampling Plans” (Efron 1982).

SASS variances can be calculated using the replicates of the full sample that are available in the data files with software such as WesVarPC. For examples of other software that support BRR, see *Introduction to Variance Estimation* (Wolter 1985).

Average Design Effects. Design effects (*Deffs*) measure the impact of the complex sample design on the accuracy of a sample estimate, in comparison to the alternative simple random sample design. For the 1990–91 SASS, an average design effect was derived for groups of statistics and, within each group, for a set of subpopulations. Standard errors for 1990–91 and 1993-94 SASS statistics of various groups for various subpopulations can then be calculated approximately from the standard errors based on the simple random sample (using SAS or SPSS) in conjunction with the average design effects provided. For example, for the 1990–91 SASS, average design effects for selected variables in the School Survey are 1.60 (public sector) and 1.36 (private sector); in the Principal Survey, 4.40 (public sector) and 4.02 (private sector); and in the Teacher Survey, 3.75 (public sector) and 2.52 (private sector). Examples illustrating the use of SASS average design effect tables are provided in *Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS), Volume I, User's Manual* (Salvucci and Weng 1995).

Generalized Variance Functions (GVFs). GVF tables were developed for use in the calculation of standard errors of totals, averages, and proportions of interest in the 1990–91 SASS components. The 1990–91 GVFs can be used for the 1993–94 SASS because no major design changes were adopted between 1990–91 and 1993–94. Note that the GVF approach, unlike the design effect approach described above, involves no need to calculate the simple random sample variance estimates. Examples illustrating the use of the GVF tables are provided in *Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS), Volume I, User's Manual* (Salvucci and Weng 1995).

Nonsampling Error

Coverage Error. SASS surveys are subject to any coverage error present in the CCD and PSS data files, which serve as their principal sampling frames. The report *Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary Education Agency Universe Survey* (Owens 1997) found that overall coverage in the 1994–95 CCD Local Education Agency Universe Survey was 96.2 percent (in a comparison to state education directories). "Regular" agencies—those traditionally responsible for providing public education—had almost total coverage in the 1994–95 agency universe survey. Most coverage discrepancies were attributed to nontraditional agencies that provide special education, vocational education, and other services. However, there is potential for undercoverage bias associated with the absence of schools built between the time when the sampling frame is constructed and the time of the SASS survey administration. Further research on coverage can be found in "Evaluating the Coverage of the U.S. National Center for Education Statistics' Public Elementary/Secondary School Frame" (Hamann 2000) and "Evaluating the Coverage of the U.S. National Center for Education Statistics' Public and Private School Frames Using Data from the National Assessment of Educational Progress" (Lee, Burke, and Rust 2000).

A capture-recapture methodology was used to estimate the number of private schools in the United States and to estimate the coverage of private schools in the 1999–2000 PSS; the study found that the PSS school coverage rate is equal to 97 percent. (See chapter 2 for a description of the CCD and chapter 3 for a description of the PSS.)

Nonresponse Error.
Unit nonresponse. The weighted unit response rates for public schools have been higher than the weighted unit response rates for private schools in all six rounds of SASS. (See table 3 for response rates from selected years.) For more information on the analysis of nonresponse rates, refer to *An Analysis of Total Nonesponse in the 1993–94 Schools and Staffing Survey (SASS)* (Monaco et al. 1997) and *An Exploratory Analysis of Response Rates in the 1990–91 Schools and Staffing Survey (SASS)* (Scheuren et al. 1996).

Item Nonresponse. For the 2007–08 SASS, the weighted item response rates for the individual surveys were as follows: 52 to 100 percent for public school districts; 71 to 100 percent for public schools; 49 to 100 percent for private schools; 65 to 100 percent for BIE schools; 76 to 100 percent for public school principals; 86 to 100 percent for private school principals; and 61 to 100 percent for BIE school principals. For teachers, the ranges of item response rates were as follows: 44 to 100 percent for public school teachers; 64 to 100 percent for private school teachers; and 0 to 100 percent for BIE teachers. Item response rates for public school library media centers and BIE school library media centers ranged from 84 to 100 percent and 71 to 100 percent, respectively.

Measurement Error. Results reported in *An Analysis of Total Response in the 1993–94 Schools and Staffing Survey (SASS)* (Monaco et al. 1997) support the contention that, without follow-up to mail surveys, nonresponse error would be much greater than it is and that the validity and reliability of the data would be considerably reduced. However, because of the substantial amount of telephone follow-up, there is

Table 3. Summary of weighted response rates for selected SASS questionnaires

Questionnaire	1993–94	1999–2000	2003–04	2007–08
School District Survey	93.9	88.6	82.9	87.8
Public Principal Survey	96.6	90.0	82.2	79.4
Public School Survey	92.3	88.5	80.8	80.4
Public Teacher Survey[1]	83.8	83.1	75.7	84.0
Private Principal Survey	87.6	84.8	74.9	72.2
Private School Survey	83.2	79.8	75.9	75.9
Private Teacher Survey[1]	72.9	77.2	70.4	77.5
BIE Principal Survey	98.7	93.3	90.7	79.2
BIE School Survey	99.3	96.7	89.5	77.1
BIE Teacher Survey	86.5	87.4	86.3	81.8

[1]The overall teacher response rates are the percentage of teachers responding in schools that provided teacher lists for sampling.
SOURCE: Aritomi, P., and Coopersmith, J. (2009). *Characteristics of Public School Districts in the United States: Results From the 2007-08 Schools and Staffing Survey* (NCES 2009-320). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Gruber, K.J., Rohr, C.L., and Fondelier, S.E. (1996). *1993–94 Schools and Staffing Survey: Data File User's Manual* (NCES 96-142). National Center for Education Statistics, U.S. Department of Education. Washington, DC; Tourkin, S.C., Pugh, K.W., Fondelier, S.E., Parmer, R.J., Cole, C., Jackson, B., Warner, T., and Weant, G. (2004). *1999–2000 Schools and Staffing Survey (SASS) Data File User's Manual* (NCES 2004-303). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Tourkin, S.C., Warner, T., Parmer, R., Cole, C., Jackson, B., Zukerberg, A., Cox, S., and Soderborg, A. (2007). *Documentation for the 2003–04 Schools and Staffing Survey* (NCES 2007-337). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

concern about possible bias due to differences in the mode of survey collection. Other possible sources of measurement error include long, complex instructions that respondents either do not read or do not understand, navigation problems related to the format of the questionnaires, and definitional and classification problems. See also *Measurement Error Studies at the National Center for Education Statistics* (Salvucci et al. 1997).

Several NCES working papers also address measurement error. Reports on the 1993–94 SASS include *Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey* (Jenkins and Von Thurn 1996); *Further Cognitive Research on the Schools and Staffing Survey (SASS)* (Zukerberg and Lee 1997); *Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993–94 School Year* (Jenkins 1997); and *Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report* (Bushery, Schreiner, and Sebron 1998). Reports on the 1991–92 SASS include the *1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report* (Royce 1994) and *The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation* (Jenkins and Wetzel 1995).

6. CONTACT INFORMATION

For content information on SASS, contact:

Kerry Gruber
Phone: (202) 502–7349
E-mail: kerry.gruber@ed.gov
SASS e-mail: sassdata@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Broene, P. (1991a). *User's Manual: 1987–88 Schools and Staffing Survey - Public School Administrator Questionnaire Data* (NCES 91-137). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Broene, P. (1991b.) *User's Manual: 1987–88 Schools and Staffing Survey - Public School Data* (NCES 91-136). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Gruber, K.J., Rohr, C.L., and Fondelier, S.E. (1994). *1990–91 Schools and Staffing Survey: Data File User's Manual Volume I: Survey Documentation* (NCES 93-144-I). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Gruber, K.J., Rohr, C.L., and Fondelier, S.E. (1996). *1993–94 Schools and Staffing Survey: Data File User's Manual, Volume I: Survey Documentation* (NCES 96-142). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Rouk, Ü., Weiner, L., and Riley, D. (1999). *What Users Say About Schools and Staffing Survey Publications* (NCES Working Paper 1999-10). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Tourkin, S.C., Pugh, K.W., Fondelier, S.E., Parmer, R.J., Cole, C., Jackson, B., Warner, T., and Weant, G. (2004). *1999–2000 Schools and Staffing Survey (SASS) Data File User's Manual* (NCES 2004-303). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourkin, S., Thomas, T., Swaim, N., Cox, S., Parmer, R., Jackson, B., Cole, C., and Zhang, B. (2010). *Documentation for the 2007-08 Schools and Staffing Survey* (NCES 2010-332). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourkin, S.C., Warner, T., Parmer, R., Cole, C., Jackson, B., Zukerburg, A., Cox, S., and Soderborg, A. (2007). *Documentation for the 2003–04 Schools and Staffing Survey* (NCES 2007-337). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Uses of Data

Ingersoll, R.M. (1998). *An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey* (NCES Working Paper 98-18). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

McGrath, D.J., and Luekens, M.T. (2000). *A Research Agenda for the 1999–2000 Schools and Staffing Survey* (NCES Working Paper 2000-10). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Mullens, J.E., and Kasprzyk, D. (1997). *The Schools and Staffing Survey: Recommendations for the Future* (NCES 97-596). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Wiley, S.D., and Reynolds, K.A. (1999). *Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results* (NCES Working Paper 1999-02). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Survey Design

Abramson, R., Cole, C., Fondelier, S., Jackson, B., Parmer, R., and Kaufman, S. (1996). *1993–94 Schools and Staffing Survey: Sample Design and Estimation* (NCES 96-089). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Baker, D.P., Levine, R., Han, M., and Garet, M. (1998). *A Feasibility Study of Longitudinal Design for Schools and Staffing Survey* (NCES Working Paper 98-16). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingersoll, R.M. (1995). *An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey* (NCES Working Paper 95-18). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingersoll, R.M. (1996). *National Assessments of Teacher Quality* (NCES Working Paper 96-24). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Isaacs, J.B., Best, C.M., Cullen, A.D., Garet, M.S., and Sherman, J.D. (1998). *Collection of Public School Expenditure Data: Development of a Questionnaire* (NCES Working Paper 98-01). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Isaacs, J.B., Garet, M.S., Sherman, J.D., Cullen, A., and Phelps, R. (1999). *Collection of Resource and Expenditure Data on the Schools and Staffing Survey* (NCES Working Paper 1999-07). National

Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kaufman, S. (1991). *1988 Schools and Staffing Survey Sample Design and Estimation* (NCES 91-127). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kaufman, S., and Huang, H. (1993). *1990–91 Schools and Staffing Survey: Sample Design and Estimation* (NCES 93-449). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Levine, R.E., Chambers, J.G., Dueñas, I.E., and Hikido, C.S. (1997). *Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey* (NCES Working Paper 97-42). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Rollefson, M. (1998). *The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper* (NCES Working Paper 98-08). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Bushery, J., Schreiner, I., and Sebron, G. (1998). *Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report* (NCES Working Paper 98-02). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Efron, B. (1982). The Jackknife, the Bootstrap, and Other Resampling Plans. *Society of Industrial and Applied Mathematics CBMS-NSF Monographs*, 38.

Hamann, T.A. (2000). Evaluating the Coverage of the U.S. National Center for Education Statistics' Public Elementary/Secondary School Frame. In *The Second International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions* (pp. 79–88). Arlington, VA: American Statistical Association.

Jabine, T. (1994). *Quality Profile for SASS: Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS)* (NCES 94-340). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jenkins, C.R. (1997). *Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993–94 School Year* (NCES Working Paper 97-10). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jenkins, C.R., and Von Thurn, D. (1996). *Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey* (NCES Working Paper 96-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jenkins, C.R., and Wetzel, A. (1995). *The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation* (NCES Working Paper 95-10). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kalton, G., Winglee, M., Krawchuk, S., and Levine, D. (2000). *Quality Profile for SASS Rounds 1–3: 1987–1995, Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS)* (NCES 2000-308). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Lee, H., Burke, J., and Rust, K. (2000). Evaluating the Coverage of the U.S. National Center for Education Statistics' Public and Private School Frames Using Data From the National Assessment of Educational Progress. In *The Second International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions* (pp. 89–98). Arlington, VA: American Statistical Association.

Monaco, D., Salvucci, S., Zhang, F., and Hu, M. (1997). *An Analysis of Total Nonresponse in the 1993–94 Schools and Staffing Survey (SASS)* (NCES 98-243). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Owens, S. (1997). *Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary Education Agency Universe Survey* (NCES 97-505). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Royce, D. (1994). *1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report* (NCES 94-03). National Center for Education

Statistics, U.S. Department of Education. Washington, DC.

Salvucci, S., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). National Center for Education Statistics, Sciences, U.S. Department of Education. Washington, DC.

Salvucci, S., Holt, A., and Moonesingle, R. (1995). *Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS): Volume II, Technical Report* (NCES 95-342-II). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Salvucci, S., and Weng, S. (1995). *Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Surveys (SASS) Volume I—User's Manual* (NCES 95-342-I). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Scheuren, F., Monaco, D., Zhang, F., Ikosi, G., and Chang, M. (1996). *An Exploratory Analysis of Response Rates in the 1990–91 Schools and Staffing Survey (SASS)* (NCES 96-338). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Sitter, R. (1990). Comparing Three Bootstrap Methods for Survey Data. *Technical Report Series of the Laboratory for Research in Statistics and Probability*. Carleton University.

U.S. Department of Education, National Center for Education Statistics. (1994). A Bootstrap Variance Estimator for the Schools and Staffing Survey; Balanced Half-Sample Replication With Aggregation Units. In *Schools and Staffing Survey (SASS), Papers Presented at the Meetings of the American Statistical Association* (NCES Working Paper 94-01). Washington, DC: Author.

U.S. Department of Education, National Center for Education Statistics. (1996). Properties of the Schools and Staffing Survey Bootstrap Variance Estimator. In *Schools and Staffing Survey (SASS): 1995 Selected Papers Presented at the 1995 Meeting of the American Statistical Association* (NCES Working Paper 96-02). Washington, DC: Author.

U.S. Department of Education, National Center for Education Statistics. (2000). A Bootstrap Variance Estimator for Systematic PPS Sampling. In *Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings* (NCES Working Paper 2000-04). Washington, DC: Author.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Zukerberg, A., and Lee, M. (1997). *Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form* (NCES Working Paper 97-23). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 5: SASS Teacher Follow-up Survey (TFS)

1. OVERVIEW

The SASS Teacher Follow-up Survey (TFS) is a follow-up survey of elementary and secondary school teachers who participated in the Schools and Staffing Survey (SASS) (see chapter 4 for details on SASS). TFS is conducted for the National Center for Education Statistics (NCES) by the U.S. Census Bureau in the school year following the SASS data collection. TFS consists of a subsample of teachers who left teaching within the year after the SASS was administered and a subsample of those who continued teaching, including those who remained in the same school as in the previous year and those who changed schools.

TEACHER FOLLOW-UP SURVEY OF SCHOOL TEACHERS

TFS collects data on:

- Stayers
- Movers
- Leavers

Purpose

To measure the attrition rate for teachers, examine the characteristics of teachers who stay in the teaching profession and those who leave, obtain activity or occupational data for those who leave the position of a K–12 teacher, obtain current teaching assignment information for those who are still teaching, and collect data on attitudes about the teaching profession in general and job satisfaction in particular. TFS is designed to support estimates of public elementary, secondary, and combined school teachers and private school teachers at the national level.

Components

TFS is composed of two questionnaires: the Former Teacher Questionnaire, which collects information from sampled teachers who leave the K–12 teaching profession within the year after SASS; and the Current Teacher Questionnaire, which collects information from sampled teachers who currently teach students in any of grades prekindergarten through 12. Eligible survey respondents are teachers in public and private elementary and secondary schools in the 50 states and the District of Columbia.

Former Teacher Questionnaire. This questionnaire collects information from former teachers on their current occupation, primary activity, plans to remain in their current position, plans for further education, plans for returning to teaching, reasons for leaving teaching, possible areas of satisfaction or dissatisfaction with teaching, salary, marital status, number of children, and reasons for retirement, as well as any other information that may be related to attrition.

Current Teacher Questionnaire. This questionnaire obtains information from current teachers, including teachers who continued to teach in the same school as in the previous year and those who changed schools. It collects information on occupational status (full time, part time), primary teaching assignment by field, teaching certificate, level of students taught, areas of satisfaction or dissatisfaction, new degrees earned or pursued, expected duration in teaching, marital status, number of children, academic year base salary, time spent performing school related tasks, and effectiveness of the school administration. If the teacher is teaching in a

different school than during the SASS administration, the questionnaire obtains information on the teacher's reasons for leaving the previous school.

Periodicity

TFS is a follow-up of selected teachers from the SASS teacher surveys and is conducted during the school year following the SASS administration. It was conducted in the 1988–89, 1991–92, 1994–95, 2000–01, and 2004–05 school years (after the 1987–88, 1990–91, 1993–94, 1999–2000, and 2003–04 administrations of SASS, respectively). The most recent survey was conducted in the 2008–09 school year, collecting data from a subsample of teachers who participated in the 2007–08 SASS.

2. USES OF DATA

Data from TFS are used for a variety of purposes by Congress, state education departments, federal agencies, private school associations, teacher associations, and educational organizations. TFS can be used to research issues related to teacher turnover. Leavers, movers, and stayers can be profiled and compared in terms of teaching qualifications, working conditions, attitudes toward teaching, job satisfaction, salaries, benefits, and other incentives and disincentives for remaining in or leaving the teaching profession. TFS also provides a measure of national teacher attrition in the various fields and updates information on the education, other training, and career paths of teachers. In addition, sampled teachers can be linked to SASS data to determine relationships between local district and school policies and practices, teacher characteristics, and teacher attrition and retention.

3. KEY CONCEPTS

Key Terms

Some of the key terms used in TFS are described below. For descriptions of other terms, see "Appendix A. Key Terms for TFS" in *Documentation for the 2008–09 Teacher Follow-up Survey* (forthcoming).

Leavers. Teachers who left the teaching profession or teachers who were no longer teaching in any of grades pre-K–12 in the school year after the SASS administration (includes teachers whose status changed to short-term substitute, student teacher, or teacher aide).

Movers. Teachers who were still teaching in the school year after the SASS administration, but had moved to a different school.

Stayers. Teachers who were teaching in the same school in the year after the SASS administration as in the year of the SASS administration.

Itinerant teacher. An individual who teaches at more than one school; for example, a music teacher who teaches 3 days per week at one school and 2 days per week at another.

4. SURVEY DESIGN

Target Population

The target population is the universe of elementary and secondary school teachers who teach in public and private schools in the 50 states and the District of Columbia, in schools that had any of grades K–12 during the school year of the last SASS administration. This population is divided into two components: those who left teaching after that school year (former teachers) and those who continued teaching (current teachers).

The TFS sample of teachers includes those who left the position of a K–12 teacher in the year after SASS (leavers). It also includes those who continued to teach students in any of grades pre-K–12 or in comparable ungraded levels, including teachers who remained in the same school as in the previous year (stayers) and those who changed schools (movers). Prekindergarten is included so that sampled teachers who change assignments from teaching students in any of grades K–12 to teaching only prekindergarten students would not be considered leavers.

In SASS, the sampling frame for public schools is an adjusted version of the NCES Common Core of Data (CCD), and the sampling frame for private schools is a modified version of the NCES Private School Universe Survey (PSS). The sampling frame for the SASS teacher questionnaire consists of lists of teachers provided by schools in the SASS sample. A teacher is defined as a staff member who taught a regularly scheduled class to students in any of grades K–12 or comparable ungraded levels.

Sample Design

TFS surveys a sample of teachers who completed interviews in the previous year's SASS. The TFS sample is a stratified sample that is allocated to allow comparisons of teachers by five variables: status (stayers, movers, leavers, and unknown); school type (traditional public, public charter, and private); experience (new and experienced); grade level (elementary, middle, and secondary); and race/ethnicity (White, non-Hispanic, Black, Hispanic, and all other races/ethnicities). In the 2008–09 TFS administration, all responding SASS teachers in public schools who indicated that their first year of teaching was 2007 or 2008 were included in the sample. All other SASS responding teachers were stratified by the five variables in the following order: school type, teacher status, experience, teacher's grade level, and race/ethnicity.

Within each TFS stratum, teachers with completed interviews in SASS are sorted by a measure of size (the SASS teacher initial basic weight, which is the inverse of the probability of selection prior to any corrections identified during data collection), main subject taught as reported by the teacher in SASS (i.e., special education, general elementary, mathematics, science, English/language arts, social studies, vocational/technical, and other), Census region, SASS private school affiliation stratum (for private school teachers only), school locale (based on the 1990 Census geography), school enrollment, and SASS teacher control number.

After teachers are sorted using the above variables, they are selected within each stratum using a systematic probability proportional to size (PPS) sampling procedure. Any teacher with a measure of size greater than the sampling interval is included in the sample with certainty (i.e., automatically included). Since TFS selection probabilities are not conditioned on anything, the selected sample sizes equal the allocated sample size.

The 2008–09 TFS sample consisted of about 5,500 teachers out of the 57,000 public and private school teachers who participated in the 2007–08 SASS. (See chapter 4 for information on the SASS sample design.)

Data Collection and Processing

The 2008–09 TFS data collection was an online collection, followed by e-mail and telephone reminders, a hard-copy mailing, and telephone follow-up. The U.S. Census Bureau is the data collection agent.

Reference dates. Most data items refer to teacher status at the time of questionnaire completion. Some items refer to the past school year, the past 12 months, or the next school year.

Data collection. In the fall of the year of the survey administration, the Census Bureau mails a Teacher Status Form to each school that had at least one teacher who participated in the previous year's SASS. On this form, the school principal (or other knowledgeable staff member) is asked to report the current occupational status of each teacher listed by indicating whether that teacher (1) is still at the school in a teaching or nonteaching capacity; or (2) has left the school to teach elsewhere or to enter a nonteaching occupation. If school staff indicates that a sample teacher has moved, and the teacher did not provide contact information on his or her SASS questionnaire, the Census Bureau tries to obtain the correct home address from the U.S. Postal Service.

For the 2008–09 TFS, the link to the user IDs and passwords for access to the online questionnaire were mailed to selected SASS teachers in early February 2009. The letters were mailed to home addresses, where available; otherwise, they were mailed to the sample teacher's school as listed in the previous SASS administration.

In March 2009, Census interviewers began calling sampled teachers who had not yet completed the survey. If the interviewers were unable to contact a sampled teacher through a contact person or through directory assistance, they called the sampled teacher's school to obtain information about his or her current address or employer. Interviewers used the same online instrument to collect the data as was used by the sampled teachers to complete the survey. Teachers who had not completed the online instrument as of April 2009 were sent a hard-copy version of the questionnaire.

Editing. Surveys undergo several stages of editing. TFS data that were provided on hard-copy versions of questionnaires are converted from paper to electronic format using manual data keying. All keyed entries are 100 percent verified by the keying staff, meaning that each field is keyed twice and the results are compared automatically for discrepancies and, subsequently, verified. All survey data are then reformatted into SAS datasets in order to begin the extensive preliminary data review process. During this stage, analysts split the TFS data into two files: a former teacher file (for leavers) and a current teacher file (for stayers and movers).

The next step is to make a preliminary determination of each case's interview status recode (ISR) value; that is, whether it is an interview, a noninterview, or out-of-scope for the survey. Records classified as interviews are submitted to a series of computer edits: range checks, consistency edits, and blanking edits. Next, the records undergo a final edit to determine whether the case is eligible to be included in the survey and, if so, whether sufficient data have been collected for the case to be classified as a completed interview. A final ISR value is then assigned to each case as a result of this edit.

Estimation Methods

Estimates from TFS sample data are produced using weighting and imputation procedures.

Weighting. The general purpose of weighting is to scale up the sample estimates to represent the target survey population. In TFS, the steps for weighting types of respondents are similar to those used for SASS. For TFS, a base weight (the inverse of the sampled teacher's probability of selection) is used as the starting point. Then, a weighting adjustment is applied that reflects the impact of the SASS teacher weighting procedure. Next, a nonresponse adjustment factor is calculated and applied using information known about the respondents from the sampling frame data. Finally, a ratio adjustment factor is calculated and applied to the sample to adjust the sample totals to frame totals in order to reduce sampling variability. The product of these factors is the final weight for TFS.

Imputation. In all administrations of TFS, all items missing values are imputed for records classified as interviews. In order to fill these items with data, questionnaires are put through three independent stages of imputation. The first stage involves using items from the same TFS questionnaire or items from the corresponding SASS school or teacher questionnaire to impute the missing data. In the second stage, any remaining unanswered items are imputed using "hot-deck" imputation (in which donor records are established and used to impute data). In the third and final stage, any remaining unanswered items are imputed clerically by Census Bureau analysts. The third stage is necessary when there is no available donor or the value imputed by computer is inconsistent with values in other items.

Future Plans

SASS is now conducted on a 4-year cycle, with the next collection planned for the 2011–12 school year. TFS is also conducted on a 4-year cycle (in the school year following the SASS administration). The next TFS administration is scheduled for the 2012–13 school year.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Because the TFS sample is a subsample of the SASS teacher sample, the SASS teacher replicate weights are used to derive the TFS replicate weights. (See the discussion of sampling error and variance estimation for SASS in chapter 4.) The base weight for each TFS teacher is multiplied by each of the SASS replicate weights divided by the SASS teacher full-sample base weight for that teacher. To calculate the 88 replicate weights, which should be used for variance calculations, these TFS replicate basic weights are processed through the remainder of the TFS weighting system.

Nonsampling Error

Coverage error. There is a potential for bias to be introduced into TFS because the TFS frame only includes teachers who responded to SASS.

Nonresponse error.

*Unit nonresponse***.** The total weighted unit response rate in the 2008–09 TFS was 88 percent. The weighted response rate for former teachers (who completed the Former Teacher Questionnaire) was slightly lower than the weighted response rate for current teachers (who completed the Current Teacher Questionnaire) (85 vs. 88 percent, respectively).

The overall response rate represents the response rate to the survey, taking into consideration each stage of data collection. For a teacher to be eligible for TFS, it was necessary to have received the Teacher Listing Form from the school during the previous year's SASS data collection, which provided a sampling frame for teachers at that school, and for the teacher to have responded to the SASS teacher questionnaire. The overall response rate (shown in Table 4) is calculated as follows: SASS Teacher Listing Form response rate x SASS teacher questionnaire response rate x TFS questionnaire response rate.

*Item nonresponse***.** Item response rates indicate the percentage of respondents who answered a given survey question or item. The weighted TFS item response rates are produced by dividing the number of sampled teachers who responded to an item by the number of sampled teachers who were eligible to

Table 4. Base-weighted response rates for SASS teacher data files and TFS data files, by sector: School years 2007-08 and 2008-09

Sector	Base-weighted 2007-08 SASS Teacher Listing Form response rate	Base-weighted 2007-08 SASS teacher data file response rate	Base-weighted 2008-09 TFS response rate		Overall 2008-09 TFS response rate	
			Current teachers	Former teachers	Current teachers	Former teachers
Total	85.9	83.3	88.3	84.7	63.2	60.6
Public[1]	86.2	84.0	88.4	84.8	64.0	61.4
Private	85.1	77.5	87.1	84.4	57.4	55.7

[1]The public sector includes teachers from traditional public and public charter schools.
NOTE: Base-weighted response rates use the inverse of the probability of selection and the sampling adjustment factor.
SOURCE: Cox, S., Parmer, T., Tourkin, S., Warner, T., and Lyter, D.M. (2007). *Documentation for the 2004–05 Teacher Follow-up Survey* (NCES 2007-349). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

answer that item, and then adjusting those rates by the final weight. In the 2008–09 TFS, the weighted item response rates for the Former TeacherQuestion-naire ranged from 75 to 100 percent. The weighted item response rates for the Current Teacher Questionnaire ranged from 74 to 100 percent. The Former Teacher Questionnaire had six items that had a weighted response rate of less than 85 percent. The Current Teacher Questionnaire had four items that had a weighted response rate of less than 85 percent.

Measurement error. Reinterviews were conducted for the purpose of measuring response variance in the 1994–95 TFS. The reinterviews were conducted through two reinterview questionnaires—one for mail cases and another for telephone cases. Each questionnaire contained a subset of questions from the original questionnaire. Seventy-eight percent of the questions evaluated displayed high response variance; only 5 percent displayed low response variance. (All but one of the 54 questions on teaching methods had moderate or high response variance.) This reinterview study again confirmed that "mark all that apply" questions tend to be problematic. See *Response Variance in the 1994–95 Teacher Follow-up Survey* (Bushery et al. 1998).

Data Comparability

Care must be taken in estimating change over time in a TFS data element, because some of the measured change may not be attributable to a change in the educational system, but due to changes in the sampling frame, questionnaire item wording, or other changes. For example, the definitions of the locale codes based on the U.S. Census were revised in 2000 and again in 2003. Changes in how schools' locales are categorized over time may account for at least some changes that are noted from previous administrations. This impacts the urbanicity variables included in the data files, which are based on the 2000 Census definitions for locale codes.

For further information on the comparability of data elements, see Appendix M in *Documentation for the 2008–09 Teacher Follow-up Survey* (forthcoming). Appendix M contains crosswalks that compare items in the 2008–09 TFS with items in the 2000–01 TFS and the 2007–08 SASS Teacher Questionnaire.

6. CONTACT INFORMATION

For content information on the TFS project, contact:

Freddie Cross
Phone: (202) 502-7489
E-mail: freddie.cross@ed.gov

Kerry Gruber
Phone: (202) 502-7349
E-mail: kerry.gruber@ed.gov

SASS e-mail: sassdata@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Cox, S., Parmer, T., Tourkin, S., Warner, T., and Lyter, D.M. (2007). *Documentation for the 2004–05 Teacher Follow-up Survey* (NCES 2007-349). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Cox, S., Parmer, T., Chambers, L., Tourkin, S., and Lyter, D. (forthcoming). *Documentation for the 2008–09 Teacher Follow-up Survey: Public-Use Version*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourkin, S.C., Warner, T., Parmer, R., Cole, C., Jackson, B., Zukerberg, A., Cox, S., and Soderborg, A. (2007). *Documentation for the 2003–04 Schools and Staffing Survey* (NCES 2007-337). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Whitener, S.D. (1994). *1991–92 SASS Teacher Follow-up Survey Data File User's Manual Public-Use Version* (NCES 94-331). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Whitener, S.D. (1994). *1991–92 SASS Teacher Follow-up Survey Data File User's Manual Restricted-Use Version* (NCES 94-478). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Whitener, S.D., Gruber, K.J., Rohr, C., and Fondelier, S. (1998). *1994–95 Teacher Follow-up Survey, Data File User's Manual Public-Use Version* (NCES 98-232). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Whitener, S.D., Gruber, K.J., Rohr, C.L., and Fondelier, S.E. (1999). *1994–95 Teacher Follow-up Survey, Data File User's Manual Restricted-Use Codebook* (NCES 1999-14). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Bushery, J.M., Schreiner, I.D., and Newman-Smith, A. (1998). *Response Variance in the 1994–95 Teacher Follow-up Survey* (NCES 98-13). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jenkins, C.R., and Wetzel, A. (1995). *The Results of the 1991–92 Teacher Follow-up Survey (TFS) Re-interview and Extensive Reconciliation* (NCES 95-10). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Mullens, J.E. (1995). *Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey* (NCES 95-15). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Salvucci, S., Walter, E., Conley, V., Fink, S., and Newman-Smith, A. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 6: National Longitudinal Study of the High School Class of 1972 (NLS:72)

1. OVERVIEW

LONGITUDINAL SAMPLE SURVEY OF THE HIGH SCHOOL SENIOR CLASS OF 1972. BASE-YEAR SURVEY AND FIVE FOLLOW-UPS, ENDING IN 1986

NLS:72 collected data from:

- Students
- School administrators
- School counselors
- Postsecondary transcripts

In response to the need for policy-relevant, time-series data on nationally representative samples of elementary and secondary students, NCES instituted the National Longitudinal Studies (NLS) Program, a continuing long-term project. The general aim of this program is to study the educational, vocational, and personal development of students at various grade levels and the personal, familial, social, institutional, and cultural factors that may affect that development. The National Longitudinal Study of the High School Class of 1972 (NLS:72) was the first in the series. The first four studies—NLS:72, the High School and Beyond Longitudinal Study (HS&B) (see chapter 7), the National Education Longitudinal Study of 1988 (NELS:88) (see chapter 8), and the Education Longitudinal Study of 2002 (ELS:2002) (see chapter 9)—cover the educational experience of youth from the 1970s into the 21st century.

NLS:72 collected comprehensive base-year data from a nationally representative sample of high school seniors in spring 1972, prior to high school graduation. Additional information about students and schools was obtained from school administrators and counselors. Over the course of the project—extending from the base-year survey in 1972 to the fifth follow-up survey in 1986—data were collected on nearly 23,000 students. A number of supplemental data collection efforts were also undertaken, including a Postsecondary Education Transcript Study (PETS) in 1984–85 and a Teaching Supplement in 1986.

Purpose
To provide information on the transition of young adults from high school through postsecondary education and into the workplace.

Components
NLS:72 collected data from students (high school seniors in 1972), school administrators, and school counselors. Data were primarily collected in a base-year and five follow-up surveys. The project also included periodic supplements completed by 1972 high school seniors and a collection of postsecondary transcripts from the colleges and universities attended by the students.

Base-year survey. The base-year survey was conducted in spring 1972 and comprised the following:

Student Questionnaire. Students reported information about their personal and family background (age, sex, race, physical handicaps, socioeconomic status [SES] of family and community); education and work experiences (school characteristics and performance; work status, performance, and satisfaction); future plans (work, education, and/or military); and aspirations, attitudes, and opinions. Students also completed a *Test Battery*—six timed aptitude tests that measured verbal and nonverbal abilities. These tests covered vocabulary, picture number, reading, letter groups,

mathematics, and mosaic comparisons. (See "Test Battery" in Section 3. Key Concepts.)

Student Record Information Form (SRIF). School administrators completed this form for each student sample member. The SRIF collected data on each student's high school curriculum, credit hours in major courses, and grade point average (and, if applicable, the student's position in ability groupings, remedial-instruction record, involvement in certain federally supported programs, and scores on standardized tests).

School Questionnaire. School administrators provided data on program and student enrollment information, such as grades covered, enrollment by grade, curricula offered, attendance records, racial/ethnic composition of school, dropout rates by sex, number of handicapped and disadvantaged students, and percentage of recent graduates in college.

Counselor Questionnaire. One or two counselors in each school provided data on their sex, race, and age; college courses in counseling and practice background; total years of counseling and years at present school; prior counseling experience with Black, Hispanic, and other race/ethnicity groups; sources of support for postsecondary education recommended to/used by students; job placement methods used; number of students assigned for counseling and number counseled per week; time spent in counseling per week; time spent with students about various problems, choices, and guidance; and time spent in various other activities (e.g., conferences with parents and teachers).

Follow-up surveys. In 1973, 1974, 1976, 1979, and 1986, NCES conducted follow-up surveys of students in the 1972 base-year sample and of students in an augmented sample selected for the first follow-up. These surveys collected information from the 1972 high school seniors on marital status; children; community characteristics; education, military service, and/or work plans; educational attainment (schools attended, grades received, credits earned, financial assistance); work history; attitudes and opinions relating to self-esteem, goals, job satisfaction, and satisfaction with school experiences; and participation in community affairs or political activities. School Questionnaires and retrospective high school data were collected during the first follow-up for sample schools and students who had not participated in the base-year survey.

Concurrently with the second follow-up, an *Activity State Questionnaire* was administered to sample members who had not provided activity information in the base-year or first follow-up surveys. Data were collected on pursuits in which the sample member was active in October of 1972 and 1973, including education, work, military service, and being a housewife, among others. Background information about the sample member's high school program and about parents' education and occupation was also requested.

During the fourth follow-up survey, a subsample of respondents was retested on a subset of the base-year Test Battery. In addition, a *Supplemental Questionnaire* was administered to respondents who had not reported certain information in previous surveys. The information asked for retrospectively covered the sample member's school and employment status from October 1972 through October 1976 and his or her license or diploma status as of October 1976. The questionnaires were tailored to the sample member's pattern of missing responses and consisted of two to four of the 11 possible sections.

The fifth follow-up survey offered the opportunity to gather information on the experiences and attitudes of sample members for whom an extensive history already existed. It differed from the previous follow-ups in that it was only sent to a subsample of the original respondents and targeted certain subgroups in the population. About 10 pages of new questions on marital history, divorce, child support, and economic relationships in families were included. The fifth follow-up also included a sequence of questions aimed at understanding the kinds of individuals who apply for and enroll in graduate management programs, as well as several questions about attitudes toward the teaching profession.

A *Teaching Supplement,* which was administered concurrently with the fifth follow-up, was a separate questionnaire that was sent to fifth follow-up respondents who indicated on the main survey form that they had teaching experience or training. The supplement focused on the qualifications, experiences, and attitudes of current and former elementary and secondary school teachers and on the qualifications of persons who had completed a degree in education or who had received certification, but had not actually taught. The supplement included items that asked about reasons for entering the teaching career, degrees and certification, actual teaching experience, allocation of time while working, pay scale, satisfaction with teaching, characteristics of the school in which the respondent taught, and professional activities. Former teachers were asked about their reasons for leaving the teaching profession and the career (if any) they pursued

afterward. Current teachers were asked about their future career plans, including how long they expected to remain in teaching. The supplement included six critical items: type of certification, certification subject(s), first year of teaching, beginning salary in the district where the respondent was currently teaching, years of experience, and grade level taught.

Postsecondary Education Transcript Study (PETS). To obtain data on coursework and credits for analysis of occupational and career outcomes, NCES requested official transcripts from all academic and vocational schools attended by the 1972 seniors since leaving high school. This study, conducted during 1984–85, collected transcripts from all postsecondary institutions reported by sample members in the first through fourth follow-up surveys. The information gathered from the transcripts included terms of attendance, fields of study, specific courses taken, and grades and credits earned. As the study covered a 12-year period, dates of attendance and term dates were recorded from each transcript received, allowing analysis over the whole period or any defined part.

Periodicity

The base-year survey was conducted in the spring of 1972, with five follow-ups in 1973, 1974, 1976, 1979, and 1986. Supplemental data collections were administered during all but the third follow-up. Postsecondary transcripts were collected in 1984–85.

2. USES OF DATA

NLS:72 is the oldest of the longitudinal studies sponsored by NCES. It is probably the richest archive ever assembled on a single generation of Americans. Young people's success in making the transition from high school or college to the workforce varies enormously for reasons only partially understood. NLS:72 data can provide information about the quality, equity, and diversity of educational opportunity and the effect of these factors on cognitive growth, individual development, and educational outcomes. It can also provide information about changes in educational and career outcomes and other transitions over time.

The Teaching Supplement data can be used to investigate policy issues related to teacher quality and retention. These data can be linked to data from prior waves of the Student Questionnaire for analysis of antecedent conditions and events that may have influenced respondents' career decisions. The data can also be merged with results from the fifth follow-up questionnaire, which included special questions related to teaching.

The history of the members of the class of 1972, from their high school years through their early 30s, is widely considered as the baseline against which the progress and achievements of subsequent cohorts are to be measured. Researchers have drawn on this archive since its inception. To date, the principal comparisons have been with the other three longitudinal studies: HS&B, NELS:88, and ELS:2002. Together, these four studies provide a particularly rich resource for examining the changes that have occurred in American education during the past 30 years. Data from these studies can be used to examine how student academic coursework, achievement, values, and aspirations have changed, or remained constant, throughout this period.

The NLS studies offer a number of possible time points for comparison. Cohorts can be compared on an *intergenerational or cross-cohort time-lag basis.* Both cross-sectional and longitudinal time-lag comparisons are possible. For example, cross-sectionally, NLS:72 seniors in 1972 can be compared to HS&B base-year seniors in 1980, NELS:88 second follow-up seniors in 1992, and ELS:2002 first follow-up seniors in 2004. Longitudinally, changes measured between the senior year and 2 years after graduation can be compared across studies. *Fixed time comparisons* are also possible; groups within each study can be compared to each other at different ages, but at the same point in time. Thus, NLS:72 seniors, HS&B seniors, and HS&B sophomores can all be compared in 1986—some 14, 6, and 4 years after each respective cohort completed high school. Finally, *longitudinal comparative analyses* of the cohorts can be performed by modeling the history of the age/grade cohorts. The possible comparison points and the considerations of content and design that may affect the comparability of data across the cohorts are discussed in *Trends Among High School Seniors, 1972–1992* (Green, Dugoni, and Ingels 1995) and *United States High School Sophomores: A Twenty-Two Year Comparison, 1980–2002* (Cahalan et al. 2006).

3. KEY CONCEPTS

A few key terms relating to NLS:72 are defined below.

Test Battery. Six cognitive tests were administered during the base year: (1) vocabulary (15 items, 5 minutes), a brief test using a synonym format; (2) picture number (30 items, 10 minutes), a test of associative memory consisting of a series of drawings of familiar objects, each paired with a number; (3) reading (20 items, 15

minutes), a test of comprehension of short passages; (4) letter groups (25 items, 15 minutes), a test of inductive reasoning that required the student to draw general concepts from sets of data or to form and try out hypotheses in a nonverbal context; (5) mathematics (25 items, 15 minutes), a quantitative comparison in which the student indicated which of two quantities was greater (or asserted their equality or the lack of sufficient data to determine which quantity was greater); and (6) mosaic comparisons (116 items, 9 minutes), a test measuring perceptual speed and accuracy through the use of items that required detection of small differences between pairs of otherwise identical mosaic, or tile-like, patterns.

Socioeconomic status (SES). A composite scale developed as a sum of standardized scales of father's education, mother's education, 1972 family income, father's occupation, and household items. The latter two underlying scales were computed from base-year Student Questionnaire responses. The other three underlying scales were derived from base-year responses as augmented by first follow-up responses and responses to a second follow-up resurvey in order to obtain this and other information from sample members who had failed to provide it previously. Each index component was first subjected to factor analysis that revealed a common factor with approximately equal weights for each component. Each of the components was then standardized, and an equally weighted combination of the five standard scores yielded the SES composite score. The data file contains both the raw score and a categorized SES score (SES Index).

4. SURVEY DESIGN

Target Population

The population of students who, in spring 1972, were 12th graders (high school seniors) in public and private schools located in the 50 states and the District of Columbia. Excluded were students in schools for the physically or mentally handicapped, students in schools for legally confined students, early (mid-year) graduates, dropouts, and individuals attending adult education classes.

Sample Design

Base-year survey. The NLS:72 sample was designed to be representative of the approximately 3 million high school seniors enrolled in more than 17,000 schools in the United States in spring 1972. The base-year sample design was a stratified, two-stage probability sample of students from all public and private schools in the 50 states and the District of Columbia that enrolled 12th graders in the 1971–72 school year. Excluded were schools for the physically or mentally handicapped and schools for legally confined students. A sample of schools was selected in the first stage. In the second stage, a random sample of 18 high school seniors was selected within each participating school.

The base-year first-stage sampling frame was constructed from computerized school files maintained by the U.S. Department of Education and the National Catholic Educational Association. The original sampling frame called for 1,200 schools; that is, 600 strata with two schools per stratum. The strata were defined based upon the following variables: type of control (public or private), geographic region, grade 12 enrollment size, geographic proximity to institutions of higher education, proportion of Black, Hispanic, and other race/ethnicity student enrollment (for public schools only), income level of the community, and degree of urbanization. Schools were selected with equal probability for all but the smallest size stratum (schools with enrollment under 300). In that stratum, schools were selected with probability proportional to enrollment. All selections were without replacement. To produce sufficient sizes for intensive study of disadvantaged students, schools in low-income areas and schools with high proportions of Black, Hispanic, and other race/ethnicity student enrollment were sampled at twice the rate used for the remaining schools. Within each stratum, four schools were selected, and then two of the four were randomly designated as the primary selections. The other two schools were retained as backup or substitute selections (for use only if one or both of the primary schools did not cooperate).

The second stage of the base-year sampling procedure consisted of first drawing a simple random sample of 18 students per school (or all students, if fewer than 18 were available) and then selecting 5 additional students (if available) as possible substitutes for nonparticipants. In both cases, the students within a school were sampled with equal probability and without replacement. Dropouts, early (mid-year) graduates, and individuals attending adult education classes were excluded from the sample. The oversampling of schools in low-income areas and schools with relatively high Black, Hispanic, and other race/ethnicity student enrollment led to oversampling of low-income and Black, Hispanic, and other race/ethnicity students.

Sample redefinitions and augmentations. At the close of the base-year survey, 1,040 schools (950 primary and 100 backup) of a targeted 1,200 schools and 26 "extra" backup schools had participated (school participation being defined as students from that school contributing SRIFs, Test Batteries, or Student Questionnaires). A backup school was termed "extra" if, ultimately, both

primary sample schools from that stratum also participated. An additional 21 primary schools indicated that they had no 1972 seniors. At this point, there remained several strata with no participating schools and many more with only one school. To reduce the effects of the large base-year school nonresponse, a resurvey activity was implemented in the summer of 1973 (prior to the first follow-up survey). An attempt was made to elicit cooperation from the 231 nonparticipating base-year primary sample schools and to obtain backup schools to fill empty or partially filled strata. The resurvey was successful in 205 of the 231 primary sample schools. Students from 36 backup schools were also included in order to obtain at least two participating schools in the first follow-up survey from each of the 600 original strata. Students from the 26 "extra" backup schools from the base-year survey were not surveyed during the first follow-up; however, students from 18 of these schools were included in the second and subsequent follow-up surveys to avoid elimination of cases with complete base-year data.

To compensate for base-year school undercoverage, samples of former 1972 high school seniors were selected for inclusion in the first and subsequent follow-ups from 16 sample augmentation schools (8 new strata); these schools were selected from those identified in 200 sample school districts canvassed to identify public schools not included in the original sampling frame. As before, 18 students per school were selected (as feasible) by simple random sample.

The number of students in the final sample from each sample school was taken as the number of students who were offered a chance to be in the sample and were eligible for the study. This included both respondents and nonrespondents, but excluded ineligible students, such as dropouts, early (mid-year) graduates, and those attending adult education classes. The final NLS:72 sample included 23,450 former 1972 high school seniors and 1,340 sample schools—1,150 participating primary schools, 21 primary schools with no 1972 seniors, 131 backup sample schools, 18 "extra" schools in which base-year student data had been completed, and 16 augmentation schools.

A subsample of 1,020 of the 14,630 eligible fourth follow-up sample members (those who had completed both a Student Questionnaire and a Test Battery in the base-year survey) was targeted for retests on a subset of the base-year Test Battery. Because a self-weighting subsample would have yielded an inadequate number of Black subsample members, a design option that oversampled Blacks was adopted. In addition to the stratification by race, the sample was controlled within strata on three factors believed to be highly correlated with retest ability scores: base-year ability, SES, and postsecondary educational achievement. The control was achieved by applying an implicit stratification procedure. Test results were obtained from 692 of those in the subsample. Additional retest data were requested for all fourth follow-up sample members who had participated in the base-year testing and who were scheduled for a personal interview. This resulted in additional test data for 1,960 individuals (50.3 percent of those defined as retest-eligible).

Fifth follow-up survey. The fifth follow-up sample was an unequal probability subsample of the 22,650 students who had participated in at least one of the five previous waves of NLS:72. The fifth follow-up retained the essential features of the initial stratified multistage design but differed from the base-year design in that the secondary sampling unit selection probabilities were unequal, whereas they were equal in the base-year design. This inequality of selection probabilities allowed oversampling of policy-relevant groups and enabled favorable cost-efficiency tradeoffs.

In general, the retention probabilities for students were inversely proportional to the initial sample selection probabilities. The exceptions were for (1) sample members with special policy relevance, who were retained with certainty or at a higher rate than other sample members; (2) persons with very small initial selection probabilities, who were retained with certainty; and (3) nonparticipants in the fourth follow-up, who were retained at a lower rate than other sample members because they were expected to be more expensive to locate and because they would be less useful for longitudinal analysis.

The subgroups of the original sample retained with certainty were (1) Hispanics who participated in the fourth follow-up survey; (2) teachers and "potential teachers" who participated in the fourth follow-up survey (a "potential teacher" was defined as a person who majored in education in college or was certified to teach or whose background was in the sciences); (3) persons with a 4-year or 5-year college degree or a more advanced degree; and (4) persons who were divorced, widowed, or separated from their spouses, or never-married parents. These groups overlapped and did not comprise distinct strata in the usual sense.

Teaching Supplement. The fifth follow-up sample included all sample members known to be teachers or potential teachers as of the fourth follow-up in 1979. To identify those sample members who had become teachers between the fourth and fifth follow-ups, a direct question was included in the fifth follow-up main questionnaire. Sample members were selected for the

Teaching Supplement sample if they indicated that they were (1) currently an elementary or secondary teacher; (2) formerly an elementary or secondary teacher; or (3) trained as an elementary or secondary teacher but never went into teaching. Of the 12,840 fifth follow-up respondents, 1,520 were eligible for the Teaching Supplement.

Postsecondary Education Transcript Study (PETS). In the first through fourth follow-up surveys, approximately 14,700 members of the NLS:72 cohort reported enrollment at one or more postsecondary institutions. An attempt was made to obtain a transcript from each school named by a respondent. Thus, no probabilistic sampling was done to define the PETS sample.

Data Collection and Processing

The base-year survey was administered through group administration. For the first four follow-up surveys, field operations began in the summer or fall of the survey year and continued through the spring of the following year; for example, the third follow-up survey data collection began in October 1976 and continued through June 1977. For the fifth follow-up survey, the data collection began in March 1986 and ended in mid-September 1986. The Educational Testing Service (ETS) administered the base-year survey; the Research Triangle Institute (RTI) carried out the first through fourth follow-up surveys; and the National Opinion Research Center (NORC) conducted the fifth follow-up survey.

Reference dates. Sample members in each of the first four follow-up surveys were asked about their family (marital status, spouse's status, number of children), location, and what they were doing with regard to work, education, and/or training during the first week of October of the survey year; fifth follow-up participants were asked the same questions for the first week of February 1986. Family income was requested for the preceding 2 years, and political and volunteer activities were requested for the past 24 months. Participants in each follow-up survey were also asked for summaries of educational and work experiences and activities for the intervening year(s) since the last survey. For the first four follow-up surveys, this information was requested as of the month of October in the intervening year(s) or sometimes overall for each year preceding the survey; fifth follow-up survey participants were asked detailed questions for up to four jobs and for attendance at up to two educational institutions since October 1979.

Data collection. Data collection instruments and procedures for the base-year survey were designed during the 1970–71 school year and were tested on a small sample of high school seniors in spring 1971. One year later, the full-scale NLS:72 study was initiated. Through an in-school group administration in the base year, each student was asked to complete a Test Battery (measuring both verbal and nonverbal aptitude) and applicable portions of a Student Questionnaire containing 104 questions distributed over 11 major sections. Students were given the option of completing the Student Questionnaire in school or taking it home and answering the questions with the assistance of their parents. In addition, school administrators at each participating school were asked to complete a School Questionnaire and an SRIF for each student in the sample. One or two counselors from each school in the sample were asked to complete a Counselor Questionnaire.

Follow-up surveys. In fall 1973, 1974, 1976, and 1979 and spring 1986, sample members (or a subsample) were again contacted. After extensive tracing to update the name and address files, follow-up questionnaires were mailed to the last known addresses of sample members whose addresses appeared sufficient and correct and who had not been removed from active status by prior refusal, reported death, or other reason. Respondents to the third through fifth follow-ups were offered small monetary incentives for completing the questionnaires. The mailouts were followed by a planned sequence of reminder postcards; additional questionnaire mailings; reminder mailgrams (for the first four follow-ups) and telephone calls; personal interviews; and, for the third to fifth follow-ups only, telephone interviews of nonrespondents. During personal interviews, the entire questionnaire was administered. During the telephone interviews conducted in the last three follow-ups, only critical items that were suitable for telephone administration were administered. In order to make survey procedures comparable, respondents were asked to keep a copy of the questionnaire in front of them for both telephone and in-person interviews.

In all follow-ups, returned questionnaire cases missing critical items were flagged during data entry, and data were retrieved by specially trained telephone interviewers. Although most questions were of the forced-choice type, coding was required for the open-ended questions on occupation, industry, postsecondary school, field of study, state where marriage and divorce occurred, and relationship. Occupational and industry codes were obtained from the U.S. Census Bureau's *Classified Index of Industries and Occupations, 1970* and *Alphabetical Index of Industries and Occupations, 1970.* These sources were used in all follow-ups. Coding of the names of postsecondary schools attended by respondents was accomplished using codes from NCES's *Education Directory, Colleges and Universities.* Field of study information was coded using classification of

instructional program (CIP) codes from NCES's *Classification of Instructional Programs*. In the fifth follow- up, for the first time, all codes were loaded into a computer program for quicker access. Coders entered a given response, and the program displayed the corresponding numerical code.

Prior to the fifth follow-up, all data were entered via direct access terminals. The fifth follow-up survey marked the first time that NLS:72 data were entered with a combination of keyed entry and optical scanning procedures. Using a computer-assisted data entry (CADE) system, operators were able to combine data entry with traditional editing procedures. All critical items and filter items (plus error-prone data like dollar amounts and numbers in general) were processed by CADE. The remaining data were optically scanned.

Teaching Supplement. Data collection procedures used for the Teaching Supplement, administered concurrently with the fifth follow-up, were similar to those used for the follow-up surveys.

Postsecondary Education Transcript Study (PETS). Packets of transcript survey materials were mailed to the postsecondary schools in July 1984, with a supplemental mailing in November 1984. Altogether, 24,430 transcripts were initially requested from 3,980 institutions for 14,760 NLS:72 sample members. Telephone follow-up of nonresponding schools began in September 1984, when transcripts had been received from about two-thirds of the schools.

After investigating several alternatives, NORC adapted its CADE system for processing postsecondary transcripts. A single member of the specially trained data preparation staff analyzed the transcript document to determine its general organization and special characteristics; abstracted standard information from the document into a common format; assigned standard numerical codes to such transcript data elements as major and minor fields of study, degrees earned, types of academic term, titles of courses taken, and grades and credits; and entered all pertinent information into a computer file. Combining these steps ensured that transcripts would be handled as internally consistent, integrated records of an individual's educational activity. Moreover, since all transcript processing occurred at a single station, the use of CADE reduced the number of steps at which records might be lost or misrouted or other errors introduced into the database.

Editing. For the base-year through fourth follow-up surveys, an extensive manual or machine edit of all NLS:72 data was conducted in preparing the release file for public use. Editing involved rigorous consistency checking of all routing patterns within an instrument (not just skip patterns containing "key" or critical items), as well as range checks for all items and the assignment of error or missing data codes as necessary. Checks of the hard-copy sources were required in some cases for error resolution.

Unlike the earlier surveys, all editing for the fifth follow-up was carried out as part of CADE. The machine-editing steps used in the prior follow-ups were implemented for scanned items. Since most of the filter questions in the fifth follow-up were CADE-designated items, there were few filter-dependent inconsistencies to be handled in machine editing. Validation procedures for the fifth follow-up centered on verification of data quality through item checks and verification of the method of administration for 10 percent of each telephone or personal interviewer's work. Field managers telephoned the respondent to check several items of fact and to confirm that the interviewer had conducted a personal or telephone interview or had picked up a questionnaire. No cases failed validation.

Postsecondary Education Transcript Study (PETS). The CADE system enforced predetermined range and value limitations on each field. It performed three types of error screenings: (1) a check-digit system, which disallowed entry of incorrect identification data (school codes from the Federal Interagency Committee on Education (FICE), student identification numbers, and combinations of schools and students); (2) each data field was programmed to disallow entry of illogical or otherwise incorrect data; and (3) each CIP code selected to classify a field of study or a course was confirmed by automatically displaying the CIP program name for the code next to the name (from the original CADE transcript) that the coder had entered. A sample of CADE transcripts was selected and printed from every completed data disk for supervisory review.

Estimation Methods

Data were weighted in NLS:72 to adjust for sampling and nonresponse. Various composite variables have also been computed to assist in data analyses.

Weighting. The weighting procedures used for the various NLS:72 survey data are described below.

Student files. NLS:72 student weights are based upon the inverse of the probabilities of selection through all stages of the sampling process and upon nonresponse adjustment factors computed within weighting classes. Unadjusted raw weights—the inverses of sample inclusion probabilities—were calculated for all students sampled in each survey year. These weights are a function of the school selection probabilities and the

student selection probabilities within a school. The raw weight for a case equals the raw weight for the base-year sample divided by the conditional probability of selection into that follow-up survey, given that the case was selected into the base-year sample.

Because of the various sample redefinitions and augmentations and nonresponse to the various student instruments, several sets of adjusted weights were computed for each NLS:72 survey wave. Each weight is appropriate for a particular respondent group. The general adjustment procedure used was a weighting class approach, which distributes the weights of nonrespondents to respondents who are in the same weighting class. The adjustment involves partitioning the entire student sample (respondents and nonrespondents) into weighting classes (homogeneous groups with respect to survey classification variables) and performing the adjustments within weighting class. Adjusted weights for nonrespondents are set to 0, and their adjusted weights are distributed to respondents proportionally to the respondents' unadjusted weights. Differential response rates for students in different weighting classes are reflected in the adjustment, and the weight total within each weighting class (and thus for the sample as a whole) is maintained.

The weighting class cells were defined by cross-classifying cases by several variables. For the first through fourth follow-up surveys, the weighting class cells were sex, race, high school program, high school grade point average, and parents' education. For the fifth follow-up survey, the weighting class cells were similar except that postsecondary education attendance was substituted for parents' education. In some instances, cells were combined by pooling across certain weighting class cells.

The adjusted weights in the third and fourth follow-ups are applicable only to key items in these surveys (or specified combinations of these items with items from other instruments). The restriction is related to a change in data collection procedures. One or two item nonresponse adjustment factors were calculated for each of these surveys for the nonkey items that were not asked. The appropriate adjusted weight for each survey should be multiplied by its nonresponse adjustment factor to provide a new weight that is appropriate to items in that survey that are not key (or combinations of such nonkey items with items from other instruments).

Refer to the NLS:72 *Fifth Follow-Up (1986) Final Technical Report* (Sebring et al. 1987) for complete weighting procedures and a specification of available weights and appropriate variables to which the weights apply.

Teaching Supplement file. One set of weights was specifically developed to compensate for the unequal probabilities of retention in the Teaching Supplement sample and to adjust for unit nonresponse. Theoretically, the weights project to the population of high school seniors of 1972 who have taught elementary or secondary school or who were trained to teach but never went into teaching. The weighting procedures were similar to those used in the follow-up surveys and consisted of two basic steps. The first step was the calculation of a preliminary weight based on the inverse of the cumulative probabilities of selection for the Teaching Supplement. The preliminary weight for the Teaching Supplement is the fifth follow-up adjusted weight. The second step carried out the adjustment of this preliminary weight to compensate for unit nonresponse. Respondents were cross-classified into weighting cells by race, high school grades, and status as a teacher (current or former teacher, or never taught).

School file. During the sequential determination of final school sample membership (including augmentations), several school sample weights were computed. The principal purpose of the various school weights was to serve as a basis for the subsequent computation of student weights applicable to one or more of the student instruments. Only two of the eight weights computed are of direct use in analyzing school file or other school-level data. The school file sample weight is appropriate for analyzing school-level data that potentially could be supplied by all schools, including the School Questionnaire data.

The adjusted counselor weight should be used only in analyzing the responses to the Counselor Questionnaire; however, care must be exercised when analyzing these data. This questionnaire was only administered at base-year responding schools, and data were collected from either one or two counselors at each school.

Postsecondary Education Transcript Study (PETS) file. Because the PETS did not introduce any additional subsampling into the NLS:72 sample design, it was not necessary to calculate a new raw weight for this study. Instead, the raw weight for the base-year survey was used to create three adjusted weights specifically for the analysis of transcript data. They are not meant to be associated with individual transcripts, but rather with all data for a particular individual. The first weight is a simple adjustment for nonresponse to the transcript study itself, where response is defined as an eligible case having one or more coded transcript records in the data file. The other two adjusted weights account for multiple instances of nonresponse (e.g., no transcripts, no response to the fourth follow-up survey, missing data for critical items). Nonresponse

adjustments were computed as ratio adjustments within 39 separate weighting classes. Cases were assigned to each weight class based on sex, race/ethnicity, high school grades, and high school program, and within each group by whether or not only proprietary schools were attended. The final adjusted weights are the product of the raw weight for the "completed" case and the nonresponse adjustment factor for the weighting class to which the case belongs.

Imputation. The problem of missing data was resolved for certain items by supplemental data collections, the creation of composite variables, and some imputation of activity state and other variables. Most of the variables were created by pooling information from various items. For example, the activity states for 1972 and 1973 were updated with information gleaned from the Activity State Questionnaires that were administered concurrently with second follow-up operations. While some procedures for imputing missing data for activity state variables were incorporated in the steps of defining and recoding variables, two further phases of imputation procedures were implemented. The first phase involved direct logical inferences (e.g., type of school from name and address of school); the second phase involved indirect logical inferences (e.g., impute studying full time for those whose study time is unknown but who are studying and not working).

5. DATA QUALITY AND COMPARABILITY

The survey was implemented after an extensive period of planning, which included the design and field test of survey instrumentation and procedures. Any additional questions were field-tested prior to inclusion in the survey. The NLS:72 sample design and weighting procedures assured that participants' responses could be generalized to the population of interest. Quality control activities were used throughout the data collection and processing of the survey.

Sampling Error

Statistical estimates derived from NLS:72 data are subject to sampling variability. Like almost all national samples, the NLS:72 sample is not a simple random sample. Taylor Series estimation techniques were used to compute standard errors in published NLS:72 reports.

In addition to standard errors, it is often useful to report design effects and the root mean design effect for complex surveys, such as NLS:72. Results from several NLS:72 studies suggest that a straightforward multiplicative adjustment of the simple random sample standard error equation adequately estimates the actual standard error estimate for a percentage. The three generalized mean design effects for the first, second, and third follow-up surveys are, respectively, the square root of 1.39, 1.35, and 1.44. To be conservative, the highest value—the square root of 1.44—can be used as an estimate for fourth follow-up data. For the fifth follow-up, the mean design effect for the overall NLS:72 sample is 2.64. The mean design effects indicate that an estimated percentage in the NLS:72 data is—on average—more than twice as variable as the corresponding statistic from a simple random sample of the same size. The mean design effects vary across the domains from a low of 2.0 for respondents from the highest SES quartile to a high of 3.8 for Black respondents.

Nonsampling Error

The major sources of nonsampling error in NLS:72 were coverage error and nonresponse error.

Coverage error. To identify public schools not included in the original sample frame, an additional 200 school districts were contacted after the base-year survey was completed, resulting in the identification of 45 augmentation schools. To compensate for the base-year undercoverage, samples of former 1972 high school seniors from 16 of these schools were included in the first and subsequent follow-up surveys. In addition, at the end of the base-year survey, several strata had no participating schools and many more had only one school (whereas the original sample design called for two schools). To compensate for this large school nonresponse, 205 base-year noncooperating primary schools and 36 backup schools were added to the sample prior to the first follow-up survey for "resurveying" with the original design. The former 1972 high school seniors from these augmented and resurveyed schools were asked some retrospective (senior year) questions during the first follow-up survey. These individuals—who redress the school frame undercoverage bias in the base year—do not appear in the NLS:72 base-year files that would typically be employed for comparisons of high school seniors; however, the presence of some retrospective data for these individuals permits refinement of comparisons grounded in 1972 data.

Also, while every effort was made to include in the fifth follow-up all persons with teaching experience, it is

conceivable that some individuals who entered teaching late were among the 6,000 cases not included in the fifth. follow-up subsample These individuals would not have had a chance to participate in the Teaching Supplement.

Nonresponse error. Detailed rates of response to various surveys and the availability of specific data items are provided in NLS:72 *Fifth Follow-Up (1986) Final Technical Report* (Sebring et al. 1987).

Unit nonresponse. For the NLS:72 student surveys, there were two stages of sample selection and hence two types of unit nonresponse—school and student. During the base year, sample schools were asked to permit the selection of individual high school seniors for the collection of questionnaire and test data. Schools that refused to cooperate in either stage of sample selection were dropped from the sample. The bias introduced by base-year school-level refusals is of particular concern since it carried over into successive rounds of the survey. To the extent that the students in refusal schools differed from students in cooperating schools during later survey waves, the bias introduced by base-year school nonresponse persisted from one wave to the next. (Base-year school nonresponse is addressed under “Coverage error” above.)

Also, individual students at cooperating schools could fail to take part in the base-year survey. Student nonresponse would not necessarily carry over into subsequent waves since student nonrespondents in the base year remained eligible for sampling throughout the study. However, a study of third follow-up responses indicated that response to earlier survey waves was the most important predictor of response to the third follow-up.

Due to intensive data collection procedures, the response rates to the individual NLS:72 surveys were high (80 percent or better) among eligible sample members. At the conclusion of fourth follow-up activities, a total of 12,980 individuals had provided information in each of the first five survey waves (base-year and all four follow-up surveys), representing 78 percent of the 16,680 base-year respondents. As a result of the various retrospective data collection efforts, the number of individuals with some key data elements for all time points through the fourth follow-up survey is 16,450—73 percent of the 22,650 respondents who participated in at least one survey. In conjunction with the supplemental data collection efforts, this led to a high degree of sample integrity among the key longitudinal data elements.

Only sample members who had participated in at least one of the previous five waves were eligible for selection into the fifth follow-up sample. Of the 14,430 fifth follow-up sample members (excluding the deceased), 89.0 percent (unweighted) completed questionnaires in the fifth follow-up; 92.2 percent participated in at least five of the six waves; and 62.1 percent participated in all six waves. There was moderate variation in weighted nonresponse rates by region; nonresponse was greater in the West and Northeast, lower in the South, and lowest in the North Central region. The relationship between urbanization and nonresponse was about the same as for region—13 percent for rural schools, 15 percent for urban schools, and 18 percent for suburban schools. There was marked variation in nonresponse by race; Blacks showed the highest nonresponse (22.1 percent), followed closely by Hispanics (19.8 percent) and Whites (14.0 percent). Males had a higher nonresponse rate (17.3 percent) than females (13.6 percent).

In PETS, one or more transcripts were received for 91.1 percent of the 13,830 sample members reporting postsecondary school attendance since leaving high school. A single transcript was received for 55 percent of this group, two transcripts for 27 percent, and three or more transcripts for over 9 percent. At the transcript level, 87 percent of the 21,870 “in-scope” transcripts requested were supplied by the postsecondary schools (2,570 of the 24,430 transcripts initially requested could not be obtained because the school had no record of the student’s attendance). Response rates varied from a high of 93 percent for transcripts sought from public 4-year colleges and universities to a low of 55 percent from vocational and proprietary schools. The higher response rates for public and private nonvocational schools may be attributable to their typically longer period of existence and the relative permanence of their student files. Telephone follow-up calls to nonresponding schools revealed that nearly half of the vocational school transcripts requested for NLS:72 students were unavailable.

Item nonresponse. While unit nonresponse can be adjusted for by weighting, this approach is impractical for item nonresponse. Researchers should take into account that NLS:72 respondents often skipped questions incorrectly or gave unrecognizable answers. However, efforts were made to retrieve missing data for critical items by telephone, with a success rate of over 90 percent.

Most item nonresponse in NLS:72 resulted from respondents’ limited recall of past events or misinterpretation of questions and routing instructions. Many items in the student files appear to have high nonresponse rates (i.e., above 10 percent). In most instances, these items are associated with the routing, or skip, patterns in the instruments. (A routing question is one that implicitly or explicitly directs a respondent

around other questions in the instrument.) Rather conservative rules were used to label blanks as either missing (illegitimate skip—code 98) or inapplicable (legitimate skip—code 99). With the more complex routing patterns, a large section of items was sometimes coded illegitimate (code 98) due to just one inconsistency in the pattern. The data user should be careful in interpreting data coded 98 and 99 and should further examine data that lie within complex routing patterns when they are required for analysis. Similarly, data labeled as suspect during the editing stage should be reexamined and possibly reclassified for specific analytic purposes.

Measurement error. The survey data were monitored for quality of processing and evaluated to determine the extent of any problems and the sources of errors. Some examples are given below.

Study of edit failures. If the respondent failed to answer certain key items properly, the questionnaire failed an edit and the respondent was contacted by telephone. A special study of survey responses in the third follow-up was conducted to determine why so many questionnaires (over 60 percent) failed the edit process. This study concluded that (1) the majority of edit failures associated with itemized financial questions involved the respondent’s failure to supply answers to each of the requested line items; (2) items structured as “check all responses that apply” were likely to be failed by a substantial number of respondents; and (3) overall data entry errors were low (except for items requiring itemized financial information).

Review of routing patterns. Quality control, completeness, routing, and consistency indices were created for use with the student files. Routing indices, computed identically for each survey, indicate the percentage of the routing questions that were ambiguously answered by an individual for a given instrument. The first four follow-up questionnaires contained 33, 52, 67, and 61 routine patterns, respectively. In general, 56 to 68 percent of all respondents proceeded through an instrument without violating any routing patterns; about 20 to 30 percent violated 1 to 5 routing patterns; and 7 to 15 percent violated 6 to 10 patterns. In all four instruments, a small percentage (3 to 7 percent) of sample members had great difficulty with the routing patterns and violated the instructions in more than 10 different patterns.

Monitoring of data entry. For the first four follow-up surveys, direct data entry terminals were used to key the survey data. For the Supplemental Questionnaires administered in the fourth follow-up survey, data entry error rates were computed based on three keyings. After the initial keying, a random sample of the questionnaires from each batch was selected for rekeying by two additional operators. The results were within the overall error rate tolerance established for NLS:72. The variable error rate across samples and operators on the selected questionnaires was 0.00040; the estimated character error rate was 0.00023.

Data Comparability

One of the major goals of the NLS Program is to make the data sufficiently comparable to allow cross-cohort comparisons between studies (NLS:72 vs. HS&B vs. NELS:88 vs. ELS:2002), as well as comparative analyses of data across waves of the same study. Nevertheless, data users should be aware of some variations in sample design, questionnaire and test content, and data collection methods that could impact the drawing of valid comparisons.

Sample design changes. Although the general NLS:72 sample design was similar for all waves, there are some differences worth noting. The original sample design called for two schools to be surveyed from each of 600 strata; however, at the end of the base-year survey, several strata had no participants and many more had only one. As a result of a resurvey effort during the first follow-up survey, the final sample included *at least* two participating schools from each stratum. The fifth follow-up sample design differed from the base-year design in that the student selection probabilities were equal in the base-year design but unequal in the fifth follow-up.

Reporting period differences. The first four follow-ups requested data as of October of the survey year, whereas the fifth follow-up used February 1986 as the reference date.

Content changes. Due to the increased interest in event history analysis, the fifth follow-up survey collected more detailed information than did earlier surveys on the time periods during which respondents held jobs or were in school. Instead of recording one start and stop date for each school and job, up to eight time periods (or start and stop dates) were shown. To allow for maximum user flexibility, the responses were coded into pairs of start and stop dates.

Comparisons between NLS:72 student data and PETS data. There are substantial discrepancies between student-reported postsecondary attendance in the NLS:72 follow-up surveys and the evidence obtained from official school transcripts collected in the PETS. One interpretation is that NLS:72 respondents

overreported instances of postsecondary school attendance by about 10 percent (unweighted). If so, researchers analyzing postsecondary schooling using only the survey data would overestimate significantly the extent of this activity. Coding errors could offer further explanation for the discrepancies.

Comparisons among NLS:72, HS&B, NELS:88, and ELS:2002. The four NLS studies were specifically designed to facilitate comparisons with each other. At the student level, three different kinds of comparative analyses are possible. (See Section 2. Uses of Data for more detail.) The overall sample design is similar, and a core of questionnaire items is comparable across all four studies. Additionally, item response theory methods can be used to place mathematics, vocabulary, and reading scores on the same scale for 1972, 1980, 1992, and 2004 high school seniors.

However, despite the considerable similarities among NLS:72, HS&B, NELS:88, and ELS:2002, the differences in sample definition and statistical design have implications for intercohort analysis. Also, sampling error tends to be a greater problem for intercohort comparisons than for intracohort comparisons because there is sampling error each time an independent sample is drawn. In addition, a number of nonsampling errors may arise when estimating trends based on results from two or more sample surveys. For example, student response rates differ across the four NLS studies, and the characteristics of the nonrespondents may differ as well. The accuracy of intercohort comparisons may also be influenced by differences in context and question order for trend items in the various student questionnaires; differences in test format, content, and context; and other factors, such as differences in data collection and methodology. While some effort has been made to maintain trend items over time, strict test and questionnaire overlap is not considerable across the four NLS studies. More specifically, differences exist in questionnaire construction and in mode and type of survey administration. See chapter 7 (HS&B), chapter 8 (NELS:88), and chapter 9 (ELS:2002) for additional information on the comparability of the four NLS studies.

6. CONTACT INFORMATION

For content information on NLS:72, contact:

Aurora D'Amico
Phone: (202) 502-7334
E-mail: aurora.damico@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5652

7. METHODOLOGY AND EVALUATION REPORTS

General

Jones, C. (1985). *The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up Survey and High School and Beyond Third Follow-Up Survey: Field Test Report* (ED 269-465). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

McAdams, K. (1981). *National Longitudinal Study of the High School Class of 1972: An Historical Overview and Summary* (ED 217-051). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Quinn, P. (1995). *National Education Longitudinal Study of 1988: Second Follow-up Research and Development Working Papers* (NCES 94-251). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Riccobono, J.A. (1981). *National Longitudinal Study of the High School Class of 1972: Fourth Follow-Up Survey Final Methodological Report* (ED 217-052). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Sebring, P., Tourangeau, R., Spencer, B., Glusberg, M., Campbell, B., Singleton, M. (1987). *National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up (1986) Final Technical Report* (NCES 87-100). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Uses of Data

Ingels, S., and Baldridge, J. (1995). *National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors* (NCES Working Paper 95-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Survey Design

Rock, D.A., Hilton, T.L., Pollack, J.M., Ekstrom, R.B., and Goertz, M.E. (1985). *Psychometric Analysis of the NLS-72 and the High School and Beyond Test Batteries* (NCES 85-218). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Spencer, B., Sebring, P., and Campbell, B. (1987). *The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-up (1986) Sample Design Report* (CS 88-403c). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Westat. (1972). *Sample Design for the Selection of a Sample of Schools with Twelfth-Graders for a Longitudinal Study*. National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Cahalan, M.W., Ingels, S.J., Burns, L.J., Planty, M., and Daniel, B. (2006). *United States High School Sophomores: A Twenty-Two Year Comparison, 1980–2002* (NCES 2006-327). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Green, P., Dugoni, B.L., and Ingels, S.J. (1995). *Trends Among High School Seniors, 1972–1992* (NCES 95-380). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Henderson, L.B., and Allen, D.R. (1981). *NLS Data Entry Quality Control: The Fourth Follow-Up Survey* (ED 221-593). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Williams, S.R., and Folsom, R.E. (1977). *Bias Resulting From School Nonresponse: Methodology and Findings.* Washington, DC.

Wisenbaker, J.M. (1981). *Factors Associated With Edit Failure* (NCES 82-213). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Wisenbaker, J.M., and Kolstad, A.J. (1981.) *Factors Related to Third Follow-Up Survey Responses* (NCES 82-209). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 7: High School and Beyond (HS&B) Longitudinal Study

1. OVERVIEW

LONGITUDINAL SAMPLE SURVEY OF THE HIGH SCHOOL SOPHOMORE AND SENIOR CLASSES OF 1980; BASE-YEAR SURVEY AND FOUR FOLLOW-UPS, ENDING IN 1992

HS&B collects data from:

- Students and dropouts
- School administrators
- Teachers
- Parents
- High school transcripts
- Postsecondary transcripts
- Postsecondary financial aid records

The High School and Beyond (HS&B) Longitudinal Study was the second study conducted as part of NCES' National Longitudinal Studies Program. This program was established to study the educational, vocational, and personal development of young people, beginning with their elementary or high school years and following them over time as they take on adult roles and responsibilities. The HS&B included two high school cohorts—a senior cohort (the graduating class of 1980) and a sophomore cohort (the sophomore class of 1980). Students, school administrators, teachers, parents, and administrative records provided data for the study. HS&B results can be compared with the results of three other longitudinal studies—the National Longitudinal Study of the High School Class of 1972 (NLS:72), the National Education Longitudinal Study of 1988 (NELS:88), and the Education Longitudinal Study of 2002 (ELS:2002). (See chapters 6, 8, and 9, respectively, for descriptions of these studies.)

The HS&B covered more than 30,000 high school seniors and 28,000 high school sophomores. It consisted primarily of a base-year survey in 1980 and four follow-up surveys in 1982, 1984, 1986, and 1992. Record studies were also conducted to obtain key supplemental data on students. As part of the first follow-up, high school transcripts were requested for the sophomore cohort, providing information on the sophomores' course taking behavior through their 4 years of high school. Postsecondary transcripts were collected in 1984 for the senior cohort and in 1987 and 1993 for the sophomore cohort. In addition, student financial aid data were obtained from administrative records in 1984 for the senior cohort and in 1986 for the sophomore cohort. The HS&B project ended in 1993 after the completion of the fourth follow-up survey and a related transcripts study of the sophomore cohort.

Purpose

To (1) study longitudinally the given cohorts' educational, vocational, and personal development, beginning with their high school years, and the personal, familial, social, institutional, and cultural factors that may affect that development; and (2) compare the results with data from the NLS:72, NELS:88, and ELS:2002 to facilitate cross-cohort studies of American youth's schooling and socialization.

Components

The HS&B compiled data from a sample of students, parents, teachers, and school administrators in a base-year and four follow-up surveys. It also collected high school and postsecondary transcripts and administrative financial aid records. The various components are described below.

Base-Year Survey. The base-year survey was conducted in spring 1980 and comprised the following:

Student Questionnaire. Students were asked to (1) fill out a booklet, which included several items on the use of non-English languages as well as confidential identifying information; (2) complete a questionnaire that focused on their individual and family background, high school experiences, work experiences, future educational plans, future occupational goals, and plans for and ability to finance postsecondary education; and (3) take timed cognitive tests that measured verbal and quantitative abilities. The sophomore test battery included achievement measures in science, writing, and civics, while seniors were asked to respond to tests measuring abstract and nonverbal abilities.

School Questionnaire. Completed by an official in the participating school, this questionnaire collected information about enrollment, staff, educational programs, facilities and services, dropout rates, and special programs for handicapped and disadvantaged students.

Teacher Comment Checklist. At each grade level, teachers had the opportunity to answer questions about the traits and behaviors of sampled students who had been in their classes. The typical student in the sample was rated on average by four different teachers.

Parent Questionnaire. A sample of parents provided information about family attitudes, family income, employment, occupation, salary, financial planning, and how these affect postsecondary education and goals. The results included responses from the parents of about 3,600 sophomores and 3,600 seniors.

First Follow-up Survey. The first follow-up survey was conducted in spring 1982. As in the base-year survey, information was collected from students, school administrators, and parents. For the 1980 senior cohort, high school and postsecondary experiences were the main focus of the survey; seniors were asked about their school and employment experiences, family status, and attitudes and plans. For the 1980 sophomore cohort, the survey gathered information on school, family, work experiences, educational and occupational aspirations, personal values, and test scores of sample participants. A high school transcript collection was also part of the first follow-up for sophomore cohort members. (See section on Record Studies for more detail.)

Sophomores were classified by high school status as of 1982 (i.e., dropout, same school, transfer, or early graduate). Dropouts completed a *Not Currently in High School Questionnaire*, which included some questions from the regular Student Questionnaire but focused on their reasons for dropping out and its impact on their educational and career development. In addition to the regular Student Questionnaire, a *Transfer Supplement* was completed by members of the sophomore cohort who had transferred out of their base-year sample high school to another high school. This supplement gathered information on the reasons for transferring and for selecting a particular school, the length of the interruption in schooling and why it occurred, and particulars about the school itself (type, location, entrance requirements, size of student body, grades). Sophomore cohort members who graduated from high school ahead of schedule completed an *Early Graduate Supplement* in addition to the regular questionnaire. The Early Graduate Supplement documented the reasons for and circumstances of early graduation, the adjustments required to finish early, and respondents' activities compared with those of other out-of-school survey members (i.e., dropouts, 1980 seniors).

Second Follow-up Survey. This survey was conducted in spring 1984. For both the sophomore and senior cohorts, the survey collected data on the students' work experience, postsecondary schooling, earnings, periods of unemployment, and so forth. For seniors, postsecondary transcripts and financial aid records were also collected. (See section on record studies for more detail.)

Third Follow-up Survey. This survey was administered in spring 1986, using the same questionnaire for both the sophomore and senior cohorts. To maintain comparability with prior waves, many questions from earlier follow-up surveys were repeated. Respondents were asked to update background information and to provide information about their work experience, unemployment history, education and other training, family information (including marriage patterns), income, and other experiences and opinions. Financial aid records and postsecondary transcripts were collected for sophomores. (See section on Record Studies for more detail.)

Fourth Follow-up Survey. This survey was administered in spring 1992 only to the sophomore cohort. The survey sought to obtain valuable information on issues of access to, and choice of, undergraduate and graduate education institutions; persistence in obtaining educational goals; progress through the curriculum; rates of degree attainment and other assessments of educational outcomes; and rates of return to the individual and society. Additionally, a collection of postsecondary transcripts for sophomore

cohort members (i.e. members who had received their baccalaureate degrees and then went on to pursue graduate, doctoral, and first-professional degrees) took place in 1993.

Record Studies. The following record studies were conducted during the course of the HS&B project.

High School Transcript Study. In fall 1982, as part of the first follow-up, nearly 16,000 high school transcripts were collected for sophomore cohort students who were seniors in 1982. This data collection allowed the study of the course taking behavior of the members of the sophomore cohort throughout their 4 years of high school. Data included a six-digit course number for each course taken; course credit, expressed in Carnegie units (a standard of measurement that represents one credit for the completion of a 1-year course); course grade; year that course was taken; grade point average; days absent; and standardized test scores. (For more information, see chapter 29, which covers the High School Transcript Studies.)

Postsecondary Education Transcript Study. This study gathered data on students' academic histories since leaving high school. As part of the second follow-up in 1984, postsecondary transcripts were collected for the senior cohort. Transcripts were requested from all postsecondary institutions reported by senior cohort members in the first and second follow-up surveys. Transcript data included dates of attendance; fields of study; degrees earned; and the titles, grades, and credits of every course attempted at each institution.

In 1987 and again in 1993, postsecondary transcripts were collected for the sophomore cohort. The latter collection allowed information to be obtained on sophomore cohort members who had received their baccalaureate degrees and then went on to pursue graduate, doctoral, and first-professional degrees.

Student Financial Aid Records. In 1984, HS&B collected institutional financial aid records and federal records on the Guaranteed Student Loan Program and the Pell Grant Program for seniors who had indicated postsecondary attendance. Federal financial aid records were obtained for the sophomore cohort in 1986.

Periodicity

The base-year survey was conducted in 1980, with four follow-ups in 1982, 1984, 1986, and 1992 (the 1992 follow-up included only the sophomore cohort). High school transcripts were collected for the sophomore cohort in 1982. Postsecondary transcripts were collected for the senior cohort in 1984 and for the sophomore cohort in 1987 and 1993. Student financial aid records were collected for the senior cohort in 1984 and for the sophomore cohort in 1986.

2. USES OF DATA

The HS&B provides information on the educational, vocational, and personal development of young people as they move from high school into postsecondary education or the workforce and then into adult life. The initial longitudinal study (NLS:72) laid the groundwork for comparison with HS&B, while successive studies (NELS:88 and ELS:2002) provide a basis for further comparisons. NLS:72 recorded the economic and social conditions surrounding high school seniors in 1972 and, within that context, their hopes and plans; subsequently, it measured outcomes while also observing the intervening processes. Data on 1980 seniors from the HS&B base-year survey are directly comparable to NLS:72 data on 1972 seniors. With the follow-up data, trend comparisons can be made for the period 1972 to 1984. HS&B permits researchers to further monitor change by, for example, measuring the economic returns of postsecondary education for minorities and delineating the need for financial aid.

By following adolescents at an earlier age (beginning in eighth grade) and into the 21st century, NELS:88 expands the base of knowledge established in the NLS:72 and HS&B studies. NELS:88 first follow-up data provide a comparison point to high school sophomores 10 years earlier, as studied in HS&B; the second follow-up data allow trend comparisons of the high school class of 1992 with the 1980 seniors studied in the HS&B. The third follow-up allows comparisons with HS&B related to postsecondary outcomes. (Please see chapter 8 for detailed information on NELS:88.)

ELS:2002 further measures educational processes and outcomes, especially as such data pertain to student learning, predictors of dropping out, and high school effects on students' access to, and success in, postsecondary education and the workforce. Comparisons can be made between high school sophomores in 1980 and in 2002, and between high school seniors in 1980 and in 2004 (the first follow-up of ELS:2002) using the HS&B and ELS:2002 studies. (Please see chapter 9 for detailed information on ELS:2002.)

By comparing the results of the HS&B and its three related longitudinal studies, researchers can determine how plans and outcomes differ in response to changing conditions, or remain the same despite such changes.

The HS&B allows both cross-sectional and longitudinal analyses of the students who were sophomores or seniors in 1980. The data are used to address issues of educational attainment, employment, family formation, personal values, and community activities since 1980. For example, a major study on high school dropouts used HS&B data to demonstrate that a large number of dropouts return to school and earn a high school diploma or an equivalency certificate. Other examples of issues and questions that can be addressed are as follows:

- How, when, and why do students enroll in postsecondary education institutions?
- Do students who, while in high school, expect to complete the baccalaureate degree actually do so?
- How has the percentage of recent graduates from a given cohort who enter the workforce in their field changed over the past years?
- What are the long-term effects of not completing high school in the traditional way? How do employment and earnings event histories of traditional high school graduates differ from those of students who do not finish high school in the traditional manner?
- Do individuals who attend college earn more than those who do not attend college? What is the effect of student financial aid?
- What percentage of college graduates is eligible or qualified to enter a public service profession, such as teaching?
- How many college graduates enter the workforce full time in the area for which they are qualified?
- How, and in what ways, do public and private schools differ?

3. KEY CONCEPTS

Some of the key terms related to HS&B are defined below.

Cognitive Tests. Achievement tests administered to both cohorts in the base-year survey and only to the sophomore cohort in the first follow-up. For the sophomore cohort, the content in the base-year and first follow-up achievement tests was as follows: (1) vocabulary (21 items, 7 minutes), using a synonym format; (2) reading (20 items, 15 minutes), consisting of short passages (100–200 words) followed by comprehension questions and a few analysis and interpretation items; (3) mathematics (38 items, 21 minutes), in which students were asked to determine which of two quantities was greater, whether they were equal, or whether there were insufficient data to answer the question; (4) science (20 items, 10 minutes), based on science knowledge and scientific reasoning ability; (5) writing (17 items, 10 minutes), based on writing ability and knowledge of basic grammar; and (6) civics education (10 questions, 5 minutes), based on various principles of law, government, and social behavior. Seniors in the base-year survey were given a cognitive test with items in the following categories: vocabulary (27 items, 9 minutes), reading (20 items, 15 minutes), mathematics (33 items, 19 minutes), picture-number pairs (15 items, 5 minutes), mosaic comparisons (89 items, 6 minutes), visualization in three dimensions (16 items, 9 minutes), and questions about the test (5 minutes).

Course Offering and Course Taking. Course offering data were collected from the School Questionnaires filled out by school administrators; course offerings included regular and advanced placement curricula provided by the schools. Course taking data were collected in different ways for the sophomore and senior cohorts. For sophomores, official high school transcripts provided records of students' coursework. For the senior cohort, high school transcripts were not available; instead, coursework was self-reported by seniors in a series of items asking retrospectively about the courses and hours taken. Despite these differences in data collection, the listings of courses for the two cohorts were consistent, including major subjects in both regular and advanced placement curricula.

Socioeconomic Status (SES). The level of a student's SES was a composite variable, constructed from a set of variables from the base-year and first follow-up data, including father's occupation, father's education, mother's education, family income, and material possessions in the household.

4. SURVEY DESIGN

Target Population

High school students who were in the 10th or 12th grade in U.S. public and private schools in spring 1980.

Sample Design

HS&B was designed to provide nationally representative data on 10th- and 12th-grade students in the United States.

Base-Year Survey. In the base-year, students were selected using a two-stage, stratified probability sample design, with secondary schools as the first-stage units and students within schools as the second-stage units. Sampling rates were set so as to select in each stratum the number of schools needed to satisfy study design criteria regarding minimum sample sizes for certain types of schools. The following types of schools were oversampled to make the study more useful for policy analyses: public schools with a high percentage of Hispanic students; Catholic schools with a high percentage of Black, Hispanic, and other race/ethnicity students; alternative public schools; and private schools with high-achieving students. Thus, some schools had a high probability of inclusion in the sample (in some cases, equal to 1.0), while others had a low probability. The total number of schools in the sample was 1,120, selected from a frame of 24,730 schools with grades 10 or 12 or both (there was only one school sample in the base-year for both cohorts). Within each stratum, schools were selected with probabilities proportional to the estimated enrollment in their 10th and 12th grades.

Within each school, 36 seniors and 36 sophomores were randomly selected. In schools with fewer than 36 seniors or 36 sophomores, all eligible students were drawn in the sample. Students in all but the special strata were selected with approximately equal probabilities. (The students in the special strata were selected with higher probabilities.) Special efforts were made to identify sampled students who were twins or triplets so that their co-twins or co-triplets could be invited to participate in the study.

Substitution was carried out for schools that refused to participate in the survey. There was no substitution for students who refused, for students whose parents refused, or for students who were absent on survey day and makeup days.

First Follow-up Survey. The first follow-up sophomore and senior cohort samples were based on the base-year samples, retaining the essential features of a stratified multistage design. (For details see *High School and Beyond First Follow-Up (1982) Sample Design Report* [Tourangeau et al. 1983].)

For the sophomore cohort, all schools selected for the base-year sample were included in the first follow-up (except 40 schools that had no 1980 sophomores, had closed, or had merged with other schools in the sample). The sample also included 17 schools that received two or more students from base-year schools; school-level data from these institutions were eventually added to students' records as contextual information. However, these schools were not added to the existing probability sample of schools.

Sophomores still enrolled in their original base-year schools were retained with certainty since the base-year clustered design made it relatively inexpensive to resurvey and retest them. Sophomores no longer attending their original base-year schools were subsampled (i.e., dropouts, early graduates, students who transferred as individuals to a new school). Certain groups were retained with higher probabilities in order to support statistical research on such policy issues as excellence of education throughout society, access to postsecondary education, and transition from school to the labor force.

Students who transferred as a class to a different school were considered to be still enrolled if their original school had been a junior high school, had closed, or had merged with another school. Students who had graduated early or had transferred as individuals to other schools were treated as school leavers for the purposes of sampling. The 1980 sophomore cohort school leavers were selected with certainty or according to predesignated rates designed to produce approximately the number of completed cases needed for each of several different sample categories. School leavers who did not participate in the base-year were given a selection probability of 0.1.

For the 1980 senior cohort, students selected for the base-year sample had a known, nonzero chance of being selected for the first and all subsequent follow-up surveys. The first follow-up sample consisted of 11,995 selections from the base-year probability sample (including 11,500 of the 28,240 base-year participants and 495 of the 6,740 base-year nonparticipants). In addition, 204 nonsampled co-twins or co-triplets (who were not part of the probability sample) were included in the first follow-up sample, resulting in a total of 12,200 selections.

High School Transcript Study (1980 Sophomore Cohort). Subsequent to the first follow-up survey, high school transcripts were sought for a probability subsample of nearly 18,500 members of the 1980 sophomore cohort. The subsampling plan for the transcript study emphasized the retention of members of subgroups of special relevance for education policy analysis. Compared to the base-year and first follow-up surveys, the transcript study sample design further increased the overrepresentation of certain

race/ethnicity groups, students who attended private high schools, school dropouts, transfers, early graduates, and students whose parents completed the base-year Parent Questionnaire on financing postsecondary education. Transcripts were collected and processed for nearly 16,000 members of the sophomore cohort.

Second and Third Follow-up Surveys. The sample for the second follow-up survey of the 1980 sophomore cohort was based upon the design of the High School Transcript Study. A total of 14,830 cases were selected from the nearly 18,500 sample members retained for the transcript study. The second follow-up sample included disproportionate numbers of sample members from policy-relevant subpopulations. The sample for the senior cohort in the second follow-up consisted exactly of those sample members selected into the first follow-up sample. The senior and sophomore cohort samples for the third follow-up survey were the same as those used for the second follow-up. The third follow-up was the last survey conducted for the senior cohort. Postsecondary school transcripts were collected for all members of the senior cohort who reported attending any form of postsecondary schooling in either of the follow-up surveys. Over 7,000 individuals reported more than 11,000 instances of postsecondary school attendance.

Fourth Follow-up Survey. The fourth follow-up was composed solely of members of the sophomore cohort, and consisted exactly of those students selected into the second and third follow-up sample. For any student who had ever enrolled in postsecondary education, complete transcript information was requested from the institutions indicated by the student.

Data Collection and Processing

HS&B compiled data from six primary sources: students, school administrators, teachers, parents of selected students, high school administrative records (transcripts), and postsecondary administrative records (transcripts and financial aid). Data collection began in fall 1979 (when information from school administrators and teachers was first gathered) and ended in 1993 (when postsecondary transcripts of sophomore cohort members were collected). The National Opinion Research Center (NORC) at the University of Chicago was the contractor for the HS&B project.

Reference dates. In the base-year survey, most questions referred to the students' experience up to the time of the survey administration in spring 1980 (i.e., all 4 high school years for the senior cohort and the first 2 high school years for the sophomore cohort). In the follow-ups, most questions referred to experiences that occurred between the previous survey and the current survey. For example, the second follow-up largely covered the period between 1982 (when the first follow-up was conducted) and 1984 (when the second follow-up was conducted).

Data collection. In both the base-year and first follow-up surveys, it was necessary to secure a commitment to participate in the study from the administrator of each sampled school. For public schools, the process began by contacting the chief state school officer. Once approval was gained at the state level, contact was made with district superintendents and then with school principals. Wherever private schools were organized into an administrative hierarchy (e.g., catholic school dioceses), approval was obtained at the superior level before approaching the school principal or headmaster. The principal of each cooperating school designated a school coordinator to serve as a liaison between the NORC staff, school administrator, and selected students. The school coordinator (most often a senior guidance counselor) handled all requests for data and materials, as well as all logistical arrangements for student-level data collection on the school premises.

In the 1980 base-year survey, a single data collection method—on-campus administration—was used for both the sophomore and senior cohorts. In the first follow-up, most members of the sophomore cohort (nearly all of whom were then in the 12th grade) were resurveyed using methods similar to those of the base-year survey. However, since some of the 1980 sophomores had left school by 1982, the first follow-up survey involved on-campus administration for in-school respondents as well as off-campus group administration for school leavers (transfers, dropouts, early graduates). On-campus surveys generally were similar to those used in the base-year. Off-campus survey sessions were held afterward for school leavers in the sophomore cohort. Personal or telephone interviews were conducted with individuals who did not attend the sessions. Members of the 1980 senior cohort were surveyed primarily by mail. Nonrespondents to the mail survey (approximately 25 percent) were interviewed either in person or by telephone.

By the time of the second follow-up, the sophomore cohort was out of school. Thus, in the second (1984) and third (1986) follow-ups, data for both the sophomore and senior cohorts were collected through mailed questionnaires. Telephone and personal interviews were conducted with sample members who did not respond to the mailed survey within 2 to 3 months. Only the sophomore cohort was surveyed in

the fourth follow-up (1992). Computer-assisted telephone interviewing (CATI) was used to collect these data. The CATI program included two instruments; the first was used to locate and verify the identity of the respondent, while the second contained all of the survey questions. The average administration time for an interview was 30.6 minutes. Intensive telephone locating and field intervention procedures were used to locate respondents and conduct interviews.

Processing. Although procedures varied across survey waves, all Student Questionnaires in all waves were checked for missing critical items. Approximately 40 items in each of the main survey instruments were designated as critical or "key" items. Cases failed this edit, if a codable response was missing for any of the key items. Such cases were flagged and then routed to the data retrieval station, where staff called respondents to obtain missing information or otherwise resolve the edit failure.

The base-year procedures for data control and preparation differed significantly from those in the follow-up surveys. Since the base-year student instruments were less complex than later instruments, the completed documents were sent directly from the schools to NORC's optical scanning subcontractor for conversion to machine-readable form. The scanning computer was programmed to perform the critical item edit on Student Questionnaires and to generate listings of cases missing critical data, which were then sent to NORC for data retrieval. School and Parent Questionnaires were converted to machine-readable form by the conventional key-to-disk method at NORC.

All follow-up questionnaires were sent to NORC for receipt control and data preparation prior to being shipped to the scanning subcontractor. The second follow-up survey contained optically scannable grids for the answers to numeric questions; staff examined numeric responses for correct entry (e.g., right justification, omission of decimal points). In the third follow-up, a portion of the instrument was designed for computer-assisted data entry (CADE), while the rest was prepared for optical scanning. All major skip items and all critical items were entered by CADE. With this system, operators were able to combine data entry with the traditional editing procedures. The CADE system stepped question by question through critical and numeric items, skipping over questions that were slated for scanning and questions that were legitimately skipped because of a response to a filter question. Ranges were set for each question, preventing the accidental entry of illegitimate responses. CADE operators were also responsible for the critical item edit; those critical items that did not pass the edit were flagged for retrieval, both manually and by the CADE system. After the retrieved data were keyed, questionnaires were shipped to the scanning firm.

For the fourth follow-up, a CATI program captured the data at the time of the interview. The CATI program examined the responses to completed questions and used that information to route the interviewer to the next appropriate question. It also applied the customary edits, described below under "Editing." At the conclusion of an interview, the completed case was deposited in the database ready for analysis. There was minimal post-data entry cleaning because the interviewing module itself conducted the majority of necessary edit checking and conversion functions. A CADE system was designed to enter and code transcript data.

The first through fourth follow-ups required coding of open-ended responses on occupation and industry; postsecondary schools; major field of study for each postsecondary school; licenses, certificates, and other diplomas received; and military specialized schools, specialty, and pay grade. Coding was compatible with the coding done in NLS:72, using the same sources from NCES and the U.S. Bureau of the Census. (See chapter 6.) In the first follow-up, staff also coded open-ended questions in the Early Graduate and Transfer supplements, and transformed numeric responses to darkened ovals to facilitate optical scanning. In the third follow-up, all codes were loaded into a computer program for more efficient access. Coders typed in a given response, and the program displayed the corresponding numeric code.

In the fourth follow-up, interviewers received additional coding capabilities by temporarily exiting the CATI program and executing separate programs that assisted them in coding the open-ended responses. Data from the coding programs were automatically sent to the CATI program for inclusion in the dataset. In addition to the online coding tasks, interviewers recorded verbatim descriptions of industry and occupation. The coding scheme for industry in the fourth follow-up was a simplified version of the scheme used in previous rounds of HS&B (verbatim responses are available for more detailed coding). The coding scheme for occupation was adapted from verbatim responses received in the third follow-up. Postsecondary institutions were coded with Federal Interagency Committee on Education (FICE) codes.

Editing. In addition to the critical item edit described above, a series of edits checked the data for out-of-

range values and inconsistencies between related items. In the base-year, machine editing was limited to examining responses for out-of-range values. No interim consistency checks were performed since there was only one skip pattern.

In the first and second follow-ups, several sections of the questionnaire required respondents to follow skip instructions. Computer edits were performed to resolve inconsistencies between filter and dependent questions, detect illegal codes, and generate reports on the incidence of correctly and incorrectly answered questions. After improperly answered questions were converted to blanks, the student data were passed to another program for conversion to appropriate missing-data codes (e.g., “legitimate skip,” “refused”). Detection of out-of-range codes was completed during scanning for all questions except those permitting an open-ended response. Hand-coded data for open-ended questions (occupation, industry, institution, field of study) were matched by computer against lists of valid codes.

In the third follow-up, CADE carried out many of the steps that normally occur during machine editing. The system enforced skip patterns, range checking, and appropriate use of reserved codes—allowing operators to deal with problems or inconsistencies while they had the document in hand. For scanned items, the same machine-editing steps as those used in prior follow-ups were implemented. Since most of the filter questions were CADE-designated items, there were few filter-dependent inconsistencies to be handled in machine editing.

In the fourth follow-up, machine editing was replaced by the interactive edit capabilities of the CATI program, which tested responses for valid ranges, data field size, data type (numeric or text), and consistency with other answers or data from previous rounds. If the system detected an inconsistency due to a keying error by the interviewer, or if the respondent simply realized that he or she had made a reporting error earlier in the interview, the interviewer could go back and change the earlier response. As the new response was entered, all of the edit checks performed at the first response were again performed. The system then worked its way forward through the questionnaire using the new value in all skip instructions, consistency checks, and the like until it reached the first unanswered question, and control was then returned to the interviewer. When problems were encountered, the system could suggest prompts for the interviewer to use in eliciting a better or more complete answer.

Estimation Methods

Weighting is used to adjust for sampling and unit nonresponse.

Weighting. The weights are based on the inverse of the selection probabilities at each stage of the sample selection process and on nonresponse adjustment factors computed within weighting cells. While each wave provided weights for statistical estimation, the fourth follow-up weights can illustrate the concept of weighting. The fourth follow-up generated survey data and postsecondary transcript data. Weights were computed to account for nonresponse in both of these data collections.

First, a raw weight, unadjusted for nonresponse in any of the surveys, was calculated and included in the data file. The raw weight provided the basis for analysts to construct additional weights adjusted for the presence of any combination of data elements. However, *caution should be used if the combination of data elements results in a sample with a high proportion of missing cases.* For the survey data, two weights were computed. The first weight was computed for all fourth follow-up respondents. The second weight was computed for all fourth follow-up respondents who also participated in the base-year survey and in the first, second, and third follow-up surveys.

Two additional weights were computed to facilitate the use of the postsecondary transcript data. The collection of transcripts was based upon sophomore cohort reports of postsecondary attendance during either the third or fourth follow-up. A student may have reported attendance at more than one school. The first transcript weight was computed for students for whom at least one transcript was obtained. It is therefore possible for a student who was not a respondent in the fourth follow-up (but who was a respondent in the third follow-up) to have a nonzero value for the first transcript weight. The second transcript weight is more restrictive. It was designed to assign weights only to cases that were deemed to have complete data. Only students who responded during the fourth follow-up (and hence students for whom a complete report of postsecondary education attendance was available and for whom all requested transcripts were received) were assigned a nonzero value for the second transcript weight. For students who did not complete the fourth follow-up interview, complete transcripts may have been obtained in the 1987 transcript study, but since it was not certain that these transcripts were complete, they were given a weight of zero.

Imputation. No imputation was performed in HS&B.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Because the sample design for the HS&B cohorts involved stratification, disproportionate sampling of certain strata, and clustered probability sampling, the calculation of exact standard errors (an indication of sampling error) for survey estimates can be difficult and expensive.

Sampling error estimates for the first and second HS&B follow-ups were calculated by the method of Balanced Repeated Replication (BRR) using *BRRVAR*, a Department of Education statistical subroutine. (The BRR programs WesVar and SUREG are now available commercially.) For the base year and the third and fourth follow-ups, Taylor Series approximations were employed. More detailed discussions of the BRR and Taylor Series procedures can be found in the *High School and Beyond Third Follow-Up Sample Design Report* (Spencer et al. 1987). The Data Analysis System (DAS), included as part of the public-release file, automatically reports design-corrected Taylor Series standard errors for the tables it generates. Therefore, users of the DAS do not need to make adjustments to these estimates.

While design effects cannot be calculated for every estimate of interest to users, design effects will be similar from item to item within the same subgroup or population. Users can calculate approximate standard error estimates for items by multiplying the standard error under the simple random sample assumption by the square root of the average design effect for the population being studied.

Nonsampling Error

Nonsampling errors include coverage, nonresponse, and measurement errors.

Coverage error. Bias caused by explicit exclusion of certain groups of schools and students (e.g., special types of schools or students with disabilities or language barriers) is not addressed in HS&B technical reports. Potential coverage error in HS&B may relate to the exclusion of schools that refused to cooperate in the base-year survey. Students who refused to participate in the base-year survey were not excluded in the follow-ups. Since students were randomly selected from the sampled schools, the HS&B sample design did not entail exclusion of specified groups. (See “Sample Design,” above, in section 4.)

Nonresponse error.

Unit nonresponse. HS&B base-year student-level estimates include two components of unit nonresponse bias: bias introduced by nonresponse at the school level, and bias introduced by nonresponse on the part of students attending cooperating schools. At the school level, some schools refused to participate in the base-year survey. Substitution was carried out for refusal schools within a stratum when there were two or more schools within the stratum. The bias introduced by base-year school-level refusals is of particular concern since it carried over into successive rounds of the survey. Students attending refusal schools were not sampled during the base-year and had no chance for selection into subsequent rounds of observation. To the extent that these students differed from students from cooperating schools in later waves of the study, the bias introduced by base-year school nonresponse would persist. Student nonresponse did not carry over in this way since student nonrespondents remained eligible for sampling in later waves of the study.

In general, the lack of survey data for nonrespondents prevents the estimation of unit nonresponse bias. However, during the first follow-up, School Questionnaire data were obtained from most of the base-year refusal schools, and student data were obtained from most of the base-year student nonrespondents selected for the first follow-up sample. These data provide a basis for assessing the magnitude of unit nonresponse bias in base-year estimates.

Overall, 1,120 schools were selected in the original sample, and 811 of those schools (72 percent) participated in the survey. An additional 204 schools were drawn in a replacement sample. Student refusals and absences resulted in a weighted student completion rate of 88 percent in the base-year survey. Participation was higher in most follow-up surveys. Completion rates in the first follow-up were as follows: 94 percent for seniors; 96 percent for sophomores eligible for on-campus survey administration; and 89 percent for sophomores who had left school between the base-year and first follow-up surveys (dropouts, transfer students, and early graduates). In the second follow-up, 91 percent of senior cohort members and 92 percent of sophomore cohort members completed the survey. In the third follow-up, completion rates were 88 percent for seniors and 91 percent for sophomores. Only the sophomore cohort was surveyed in the fourth follow-up; 86 percent of the sample members participated.

As results from the fourth follow-up illustrate, student nonresponse varied by demographic and educational characteristics. Males had a slightly higher

nonresponse rate than females (a difference of slightly over 3 percent). Blacks and Hispanics showed similarly high rates of nonresponse (around 20 percent), whereas nonresponse among White students was about 10 percent. Nonresponse increased as socioeconomic status decreased. Students who were in general or vocational programs during the base-year were more likely to be nonrespondents than students in academic programs. Dropouts had higher nonresponse rates than other students. Students with lower grades and lower test scores showed higher nonresponse than students with higher grades and test scores. Students who were frequently absent from school showed higher nonresponse than students absent infrequently. Students with no postsecondary education by the time of the second follow-up had higher nonresponse than students with some postsecondary education. By selected school characteristics, the highest nonresponse rates were among students from alternative public schools, schools with large enrollments, schools in urban areas, and schools in the Northeast and West.

The patterns were similar in earlier rounds of HS&B. Nonresponse analyses conducted by NORC support the following general conclusions:

(1) The school-level bias component in HS&B estimates is small, averaging less than 2 percent for base-year and first follow-up estimates. It is probably of a similar magnitude for fourth follow-up estimates.

(2) The student-level bias component in base-year estimates is also small, averaging about 0.5 percent for percentage estimates.

(3) The student-level bias component in first, second, and third follow-up estimates is limited by the nonresponse rates, which were about three-fourths of the base-year rates.

(4) The student-level bias component in the fourth follow-up estimates is limited by the nonresponse rate, which was slightly higher than the base-year rate.

The first and second conclusions together suggest that nonresponse bias is not a major contributor to error in base-year estimates. The first and third conclusions suggest that nonresponse bias is not a major contributor to error in the first, second, and third follow-up estimates either. The first and fourth conclusions suggest that the fourth follow-up nonresponse bias might be a little greater than for the previous follow-ups, but probably not by much. Each of these conclusions must be given some qualifications. The analysis of school-level nonresponse is based on data concerning the schools, not the students attending them. The analyses of student nonresponse are based on survey data and are themselves subject to nonresponse bias. Despite these limitations, the results consistently indicate that nonresponse had a small impact on base-year and follow-up estimates.

Item nonresponse. Among students who participated in the survey, some did not complete the questionnaire or gave invalid responses to certain questions. The amount of item nonresponse varied considerably by item. For example, in the second follow-up, a very low nonresponse rate (0.1 percent) was observed for a question asking whether the respondent had attended a postsecondary institution. A much higher nonresponse rate (12.2 percent) was obtained for a question asking if the respondent had used a micro- or minicomputer in high school. Typical item nonresponse rates ranged from 3 to 4 percent.

Imputation was not used to compensate for item nonresponse in HS&B. However, an attempt was made in the fourth follow-up to reduce item nonresponse. In previous rounds, interviews were conducted by self-administered questionnaires (SAQs). Unfortunately, respondents often skipped questions incorrectly or gave unrecognizable answers. Thus, more data were missing than would have occurred through personal interviewing. In the fourth follow-up, interviewing was conducted using a CATI program. Unlike SAQs, CATI interviewing virtually eliminated missing data attributable to improperly skipped questions.

To evaluate the effectiveness of CATI interviewing, 25 items from both the third and fourth follow-up data were selected for comparison. Refusal and "don't know" responses were considered to be missing, but legitimate skips were not. For these 25 items, the overall percentage of missing items dropped from 4.36 percent in the third follow-up to 1.88 percent in the fourth follow-up.

CATI also eliminated all multiple responses and resulted in uncodable verbatim responses for only the two income variables. In addition, more was known about the missing data in the fourth follow-up. In the third follow-up, only 7.2 percent of the missing data were classified as refusals or "don't know" responses. In the fourth follow-up, 50.9 percent of the missing data were classified as refusals or "don't know" responses. The fact that most of the 25 comparisons showed a "very significant" decline in missing data supports the contention that missing data were reduced in the fourth follow-up.

Measurement error. An examination of consistency between responses to the third and fourth follow-ups provides an indication of the reliability of HS&B data.

Race/ethnicity. Race/ethnicity is one characteristic of the respondents that should not change between surveys. Overall, of the 12,310 respondents who reported their race/ethnicity on both questionnaires, 93.8 percent gave the same response in both years. However, certain race/ethnicity categories (e.g., Native American) had substantially less agreement. Only 53.4 percent of the respondents who classified themselves as Native Americans during the third follow-up classified themselves as Native Americans again during the fourth follow-up.

One explanation for these discrepancies may be the change in the method of survey administration. Unlike the third follow-up, which involved self-administered questionnaires, the fourth follow-up was conducted by telephone. The questionnaires mailed during the third follow-up had the five race/ethnicity categories listed for the respondent to see. In the fourth follow-up, respondents were simply asked over the telephone, "What is your race/ethnicity?" The interviewer coded the response. It is possible that Native Americans, Hispanics, and Asian/Pacific Islanders classified themselves as Black or White (not knowing that there was a more specific category for them to chose from), hence resulting in more Blacks and Whites in the fourth follow-up results.

Marital status. In the third follow-up, respondents were asked about their marital status in the first week of February 1986. In the fourth follow-up, respondents were asked about their marital status during and since February 1986. Although both questions asked about marital status during February 1986, respondents who had a change in marital status during the last 3 weeks of February could have given a different answer in the fourth follow-up than in the third follow-up. Overall, of the 11,850 respondents who gave their marital status in both questionnaires, 95.4 percent had answers that agreed.

Unlike the race/ethnicity question, memory and timing play an important role in matching answers for marital status. In this case, the recall period for third follow-up respondents was years shorter than the recall period for respondents in the fourth follow-up. Respondents in the third follow-up, which took place in spring 1986, were asked about a recent event. Respondents in the fourth follow-up, which was conducted in spring 1992, were asked to recall their status back in February 1986. As with the race/ethnicity question, the method of administering the question differed between rounds—namely, the question formatting had changed and the fourth follow-up used preloaded data to verify marital status.

Data Comparability

A goal of the National Longitudinal Studies Program is to allow comparative analysis of data generated in several waves of the same study as well as to enable cross-cohort comparisons with the other longitudinal studies. While the HS&B and NLS:72 studies are largely compatible, a number of variations in sample design, questionnaires, and data collection methods should be noted as a caution to data users.

Comparability within HS&B. While many data items were highly compatible across waves, the focus of the questionnaires necessarily shifted over the years in response to the changes in the cohorts' life cycle and the concerns of education policymakers. For seniors in the base-year survey and for sophomores in both the base-year and first follow-up surveys, the emphasis was on secondary schooling. In subsequent follow-ups, increasingly more items were collected dealing with postsecondary education and employment. Also, a major change in the data collection method occurred in the fourth follow-up, when CATI was introduced as the primary approach. Earlier waves used mailed questionnaires supplemented by telephone and personal interviews.

Comparability with NLS:72. The HS&B was designed to build on NLS:72 in three ways. First, the HS&B base-year survey included a 1980 cohort of high school seniors that was directly comparable to the NLS:72 cohort (1972 seniors). Replication of selected 1972 Student Questionnaire items and test items made it possible to analyze changes subsequent to 1972 and their relationship to federal education policies and programs in that period. Second, the introduction of the sophomore cohort in HS&B provided data on the many critical educational and vocational choices made between the sophomore and senior years in high school, thus permitting a fuller understanding of the secondary school experience and how it affects students. Third, HS&B expanded the NLS:72 focus by collecting data on a range of life cycle factors, such as family formation, labor force behavior, intellectual development, and social participation.

The sample design was largely similar for both HS&B and NLS:72, except that HS&B included a sophomore sample in addition to a senior sample. The questionnaires for the two studies contained a large number of identical (or similar) items dealing with secondary education and postsecondary work experience and education. The academic tests were

also highly comparable. Of the 194 test items administered to the HS&B senior cohort in the base-year, 86 percent were identical to items that had been given to NLS:72 base-year respondents. Item response theory (IRT) was used in both studies to put math, vocabulary, and reading test scores on the same scale for 1972, 1980, and 1982 seniors. With the exception of the use of CATI in the HS&B fourth follow-up, both NLS:72 and HS&B used group administration of questionnaires and tests in the earliest surveys and mailed questionnaires in the follow-ups. HS&B, however, involved more extensive efforts to supplement the mailings by telephone and personal interviews.

Comparability with NELS:88. The sample design of HS&B was also similar to that of NELS:88. In each base-year, students were selected through a two-stage stratified probability sample, with schools as the first-stage units and students within schools as the second-stage units. Because NELS:88 base-year sample members were eighth-graders in 1988, its follow-ups encompass students (both in the modal grade progression sequence and out of sequence) and dropouts. Despite similarities, however, the sample designs of the two studies differ in three major ways: (1) the NELS:88 first and second follow-ups had relatively variable, small, and unrepresentative within-school student samples, compared to the relatively uniform, large, and representative within-school student samples in the HS&B; (2) unlike the earlier study, NELS:88 did not provide a nationally representative school sample in its follow-ups; and (3) there were differences in school and subgroup sampling and oversampling strategies in the two studies. These sample differences imply differences in the respondent populations covered. (For details on NELS:88, please refer to chapter 8).

Comparability with ELS:2002. The ELS:2002 base-year and first follow-up surveys contain many data elements that are comparable to items from the HS&B. Differences in sampling rates, sample sizes, and design effects across the studies, however, affect the precision of estimation and comparability. Asian students, for example, were oversampled in ELS:2002, but not in HS&B, where their numbers were quite small. The base-year (1980) participating sample in HS&B numbered 30,030 sophomores; in contrast, 15,362 sophomores participated in the base-year of ELS:2002. Cluster sizes within schools were much larger for HS&B (on average, 30 sophomores per school) than for ELS:2002 (just over 20 sophomores per school); larger cluster sizes are better for school effects research, but carry a penalty in greater sample inefficiency. Mean design effect (a measure of sample efficiency) is also quite variable across the studies. For example, for 10^{th} grade, the design effect was 2.9 for HS&B, while a more favorable design effect of 2.4 was achieved for the ELS:2002 base-year. (For details on ELS:2002, please refer to chapter 9).

6. CONTACT INFORMATION

For content information on HS&B, contact:

Aurora M. D'Amico
Phone: (202) 502-7334
E-mail: aurora.damico@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Zahs, D., Pedlow, S., Morrissey, M., Marnell, P., and Nichols, B. (1995). *High School and Beyond Fourth Follow-Up Methodology Report* (NCES 95-426). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Uses of Data

Alt, M.N., and Bradby, D. (1999) *Procedures Guide for Transcript Studies* (NCES Working Paper 95-426). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S., and Baldridge, J. (1995). *National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors* (NCES Working Paper 95-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S., and Dowd, K. (1995). *National Education Longitudinal Study of 1988: Conducting Trend Analyses: HS&B and NELS:88 Sophomore Cohort Dropouts* (NCES Working Paper 95-07). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S., and Taylor, J. (1995). *National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data* (NCES Working Paper 95-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Survey Design

Frankel, M., Kohnke, L., Buonanno, D., and Tourangeau, R. (1981). *High School and Beyond Sample Design Report* National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jones, C., and Spencer, B.D. (1985). *High School and Beyond Second Follow-Up (1984) Sample Design Report.* National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Rock, D.A., Hilton, T.L., Pollack, J.M., Ekstrom, R.B., and Goertz, M.E. (1985). *Psychometric Analysis of the NLS-72 and the High School and Beyond Test Batteries.* National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Spencer, B.D., Sebring, P., Campbell, B., and Carroll, D. (1987). *High School and Beyond Third Follow-Up Sample Design Report* (NCES 88-402). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Tourangeau, R.E., McWilliams, H., Jones, C., Frankel, M.R., and O'Brien, F. (1983). *High School and Beyond First Follow-Up (1982) Sample Design Report.* National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Fetters, W.B., Stowe, P., and Owings, J.A. (1984). *Quality of Responses of High School Students to Questionnaire Items* (NCES 84-216). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 8: National Education Longitudinal Study of 1988 (NELS:88)

1. OVERVIEW

The National Education Longitudinal Study of 1988 (NELS:88) was the third major secondary education longitudinal survey sponsored by NCES. The first two surveys—the National Longitudinal Study of the High School Class of 1972 (NLS:72) and the High School and Beyond (HS&B) Longitudinal Study— examined the educational, vocational, and personal development of young people, beginning in high school. (See chapters 6 and 7 for descriptions of these studies.) The fourth high school longitudinal study, the Education Longitudinal Study of 2002 (ELS:2002), was designed to provide trend data about critical transitions experienced by students as they proceed through high school and into postsecondary education or their careers. (See chapter 8 for a description of this study.) NELS:88 provides new data about critical transitions experienced by students as they proceed from 8^{th} grade through high school and into postsecondary education or the workforce. It expands the knowledge base of the two previous studies by surveying adolescents at an earlier age and following them into the 21st century.

The NELS:88 base-year survey included a national probability sample of 1,052 public and private 8^{th}-grade schools, with almost 25,000 participating students across the United States. Three follow-up surveys were conducted at 2-year intervals from 1990 to 1994. In 1994 (the third follow-up), most sample members were 2 years out of high school. A fourth follow-up was conducted in 2000. In addition to surveying and testing students, NELS:88 gathered information from the parents of students, teachers, and school administrators. Furthermore, two rounds of transcript data were collected on the 8^{th}-grade cohort. High school transcripts were collected for all participants in the school-age sample, including dropouts and early graduates. Postsecondary transcripts were collected for students who reported attending a school beyond high school.

LONGITUDINAL SAMPLE SURVEY OF THE 8^{TH}-GRADE CLASS OF 1988; BASE-YEAR SURVEY AND FOUR FOLLOW-UPS THROUGH 2000

NELS:88 collected data from:

- Students and dropouts
- School administrators
- Teachers
- Parents/guardians
- High school transcripts
- High school course offerings
- High School Effectiveness Study
- Postsecondary education transcripts

Purpose
To provide trend data about critical transitions experienced by young people as they leave elementary school and progress through high school into postsecondary institutions or the workforce, and provide data for trend comparisons with results from NLS:72 and HS&B as well as later longitudinal studies, such as ELS: 2002.

Components
NELS:88 collected survey data from students, dropouts, parents, teachers, and school administrators. Supplementary information was gathered from high school transcripts and course offering data provided by the schools, a Base-Year Ineligible (BYI) Study, a Followback Study of Excluded Students (FSES), a High School Effectiveness Study (HSES), and a Postsecondary Education Transcript Study. The various components are described below.

Base-Year Survey. The base-year survey was conducted during the spring school term in 1988 and included the following:

Student Questionnaire (8th-Grade Questionnaire). Students were asked to fill out a questionnaire that included items on their home background, language use, family, opinions about themselves, plans for the future, job and chores, school life, schoolwork, and activities. Students also completed a series of curriculum-based cognitive tests in four achievement areas—reading, mathematics, science, and social studies (history/government).

Parent Questionnaire. One parent of each student completed a questionnaire requesting information about both parents' background and socioeconomic characteristics, aspirations for their children, family willingness to commit resources to their children's education, the home educational support system, and other family characteristics relevant to achievement.

Teacher Questionnaire. A Teacher Questionnaire was administered to selected 8th-grade teachers responsible for instructing sampled students in two of the four test subjects—mathematics, science, English, and social studies. The questionnaire collected information in three areas: teachers' perceptions of the sampled students' classroom performances and personal characteristics; curriculum content of the areas taught; and teachers' background and activities. Two teachers were asked to respond for each student.

School Administrator Questionnaire. Completed by an official in the participating school, this questionnaire collected information about school, student, and teacher characteristics; school policies and practices; the school's grading and testing structure; school programs and facilities; parent involvement in the school; and school climate.

First Follow-up Survey. The first follow-up survey was conducted in spring 1990. It collected information from students, teachers, and school administrators, but not parents. The student sample was freshened to be nationally representative of students enrolled in the 10th grade in spring 1990. In addition, three new components were initiated: the Dropout Questionnaire, the Base-Year Ineligible (BYI) Study, and the High School Effectiveness Study (HSES).

Students were again requested to complete a questionnaire and take cognitive tests. The Student Questionnaire collected background information and asked students about such topics as their school and home environments, participation in classes and extracurricular activities, current jobs, goals and aspirations, and opinions about themselves. Dropouts were asked similar questions in a separate *Not Currently in School Questionnaire* (or *Dropout Questionnaire*), which also requested specific information about reason(s) for leaving school and experiences in and out of school. Dropouts were also given cognitive tests when feasible.

School administrators provided information about their high schools in the School Administrator Questionnaire, and two teachers for each student completed the Teacher Questionnaire. There were different Teacher Questionnaires for English, mathematics, science, and history. The School Administrator and Teacher Questionnaires provided information about school administration, school programs and services, curriculum and instruction, and teachers' perceptions about their students' learning.

Second Follow-up Survey. The second follow-up survey, conducted in 1992, repeated all the components of the first follow-up survey and included the Parent Questionnaire. The student sample was again freshened to be nationally representative of students enrolled in the 12th grade in spring 1992. A new High School Transcript Study provided archival data on the academic experience of high school students. Students in high schools designated in the first follow-up for HSES were surveyed and tested again in both the main second follow-up survey and a separate HSES.

As in the previous waves, students were asked to complete a questionnaire and cognitive tests. The cognitive tests were designed to measure 12th-grade achievement and cognitive growth between 1988 and 1992 in mathematics, science, reading, and social studies (history/citizenship/geography). The questionnaire asked students about such topics as academic achievement; perceptions about their curricula and schools; family structures and environments; social relations; and aspirations, attitudes, and values relating to high school, occupations, and postsecondary education. The Student Questionnaire also contained an *Early Graduate Supplement*, which asked early graduates to document the reasons for and circumstances of their early graduation. Students who were first-time participants in NELS:88 completed a *New Student Supplement*, containing basic demographic items requested in the base year but not repeated in the second follow-up. First follow-up dropouts were resurveyed and retested. School administrators completed the School Administrator Questionnaire, and one mathematics or science teacher for each student completed the Teacher Questionnaire.

Third Follow-up Survey. The third follow-up survey, conducted in 1994, contained only the Student Questionnaire, which collected information mainly on

issues related to employment and postsecondary education. Specific content areas included academic achievement; perceptions and feelings about school and/or job; work experience and work-related training; application and enrollment in postsecondary education institutions; sexual behavior, marriage, and family; and values, leisure-time activities, volunteer activities, and voting behavior.

Fourth Follow-up Survey. The fourth follow-up survey, conducted in 2000, contained only the Student Questionnaire, which collected information mainly on issues of employment and postsecondary education. Specific content areas included academic achievement; perceptions and feelings about school and/or job; work experience and work-related training; application and enrollment in postsecondary education institutions; sexual behavior, marriage, and family; and values, leisure-time activities, volunteer activities, and voting behavior.

Supplemental Studies. The following supplemental studies were conducted during the course of the NELS:88 project:

Base-Year Ineligible (BYI) Study. The BYI Study was added to the first follow-up survey to ascertain the status of students who were excluded from the base-year survey due to a language barrier or physical or mental disability that precluded them from completing a questionnaire and cognitive tests. Any students found to be eligible at this time were included in the follow-up surveys.

Followback Study of Excluded Students (FSES). This study—a part of the second follow-up survey—was a continuation of the first follow-up BYI Study.

High School Transcript Study. This study collected high school transcripts during the second follow-up survey. Complete transcript records were collected for (1) students attending sampled schools in spring 1992; (2) dropouts (including those in alternative programs) and early graduates; and (3) sample members who were ineligible for any wave of the survey due to mental or physical disability or language barriers. The transcript data collected from schools included student-level data (e.g., number of days absent per school year, standardized test scores) and complete course-taking histories (e.g., information on credits earned; year and term a specific course was taken; and final grades). (For more information, see chapter 29, High School Transcript Studies.)

High School Effectiveness Study (HSES). To facilitate longitudinal analysis at the school level, a School Effects Augmentation was implemented in the first follow-up survey to provide a valid probability sample of 10th-grade schools. From the pool of NELS:88 first follow-up schools, a probability subsample of 251 urban and suburban schools in the 30 largest Metropolitan Statistical Areas was selected for the HSES; 248 of these schools were HSES participants in the first follow-up. The NELS:88 national or “core” student sample in these schools was augmented to obtain a within-school representative student sample large enough to support school effects research (i.e., the effects of school policies and practices on students). These schools and students were followed up in 1992—when the majority of the students were in 12th grade—as part of both the main NELS:88 second follow-up survey and the HSES. The HSES also provided a convenient framework for a constructed-response testing experiment in 1992. The test contained four questions that required students to derive answers from their own knowledge and experience (e.g., write an explanation, draw a diagram, solve a problem). Mathematics tests were assigned to half of the schools that were willing to commit the extra time required for such testing; the other half were assigned science tests. The second follow-up HSES was also enhanced by the collection of curriculum offerings in the Course Offerings Component. (See below.)

Course Offerings Component. This component was added to the second follow-up to provide curriculum data that can serve as a baseline for studying student outcomes. The course offerings data for this component were collected from the HSES schools. These data illuminate trends when examined in conjunction with data from the transcript studies conducted as part of the 1982 HS&B and the 1987, 1990, 1994, and 1998 National Assessment of Educational Progress (NAEP).

Postsecondary Education Transcript Study. The Postsecondary Education Transcript Study was conducted as part of the fourth follow-up survey in 2000. It targeted transcripts from all U.S. postsecondary institutions attended by NELS sample members in the fourth follow-up, excluding postsecondary information collected from foreign institutions, non-degree-granting programs, and non-credit-granting institutions. The Postsecondary Education Transcript Study supplements the postsecondary education information collected in the 1994 and 2000 follow-ups by including detailed information on types of degree programs, periods of enrollment, majors or fields of study for instructional programs, specific courses taken, grades and credits attained, and credentials earned.

Periodicity

Biennial from 1988 to 1994, a fourth follow-up was conducted in 2000. The Base-Year Ineligible Study

was conducted in 1990 as part of the first follow-up; a continuation study, the Followback Study of Excluded Students, was conducted in 1992 as part of the second follow-up. The High School Effectiveness Study was conducted in the first and second follow-ups. The High School Transcript Study was implemented in the second follow-up in 1992. The Postsecondary Education Transcript Study was conducted as part of the fourth follow-up in 2000.

2. USES OF DATA

The NELS:88 project was designed to provide trend data about critical transitions experienced by students as they leave elementary school and progress through high school and into postsecondary education or the workforce. Its longitudinal design permits the examination of changes in young people's lives and the role of school in promoting growth and positive life outcomes. The project collects policy-relevant data about educational processes and outcomes, early and late predictors of dropping out, and school effects on students' access to programs and equal opportunity to learn. These data complement and strengthen state and local efforts by furnishing new information on how school policies, teacher practices, and family involvement affect student educational outcomes (e.g., academic achievement, persistence in school, and participation in postsecondary education).

NELS:88 data can be used in three ways: in cross-sectional, longitudinal, and cross-cohort analyses (by comparing NELS:88 findings with those of NLS:72, HS&B, and ELS:2002). By following young adolescents at an earlier age (beginning in 8th grade) and into the 21st century, NELS:88 expands the base of knowledge established in the NLS:72 and HS&B studies. NELS:88 first follow-up data provide a comparison point to high school sophomores 10 years earlier, as studied in HS&B. NELS:88 second follow-up data allow trend comparisons of the high school class of 1992 with the 1972 and 1980 seniors studied in NLS:72 and HS&B, respectively. The NELS:88 third follow-up allows comparisons with NLS:72 and HS&B related to postsecondary outcomes. ELS:2002 is different from NELS:88 in that the base-year sample students are 10th-graders rather than 8th-graders. With a freshened senior sample, the ELS:2002 first follow-up supports comparisons with the NELS:88 second follow-up. The ELS:2002 first follow-up academic transcript component also offers a further opportunity for a cross-cohort comparison with the high school transcript studies of NELS:88. Together, the four studies provide measures of educational attainment in the United States and rich resources for studying the reasons for and consequences of academic success and failure.

More specifically, NELS:88 data can be used to investigate

- *transitions from elementary to secondary school*: how students are assigned to curricular programs and courses; how such assignments affect their academic performance as well as future career and postsecondary education choices;
- *academic growth over time*: family, community, school, and classroom factors that promote growth; school classroom characteristics and practices that promote learning; effects of changing family composition on academic growth;
- *features of effective schools*: school attributes associated with student academic achievement; school effects analyses;
- *the dropout process*: contextual factors associated with dropping out; movement in and out of school, including alternative high school programs;
- *the role of the school in helping the disadvantaged*: school experiences of the disadvantaged; approaches that hold the greatest potential for helping them;
- *school experiences and academic performance of language-minority students:* variation in achievement levels; bilingual education needs and experiences;
- *students' mathematics and science learning :* math and science preparation received by students; student interest in these subjects; encouragement by teachers and school to study advanced mathematics and science; and
- *transitions from high school to college and postsecondary access/choice*: planning and application behaviors of the high school class of 1992; subsequent enrollment in postsecondary institutions.

3. KEY CONCEPTS

Some of the key terms related to NELS:88 are defined below.

Cognitive Test Battery. Cognitive tests measuring student achievement in mathematics, reading, science, and social studies (history/citizenship/geography) were administered in the base year, first follow-up, and second follow-up. The contents was as follows: (1) reading (21 items, 21 minutes); (2) mathematics (40 items, 30 minutes); (3) science (25 items, 20 minutes); and (4) social studies (30 items, 14 minutes—the base-year test included history and government items; the first and second follow-up tests included history, citizenship, and geography items).

Socioeconomic Status (SES). A composite variable constructed from five questions in the Parent Questionnaire: father's education level, mother's education level, father's occupation, mother's occupation, and family income. When all parent variables were missing, student data were used to compute the SES, substituting household items (e.g., dictionary, computer, more than 50 books, washing machine, calculator) for the family income variable. There are separate SES variables derived from parent data in the base year and the second follow-up. The database also included variables for SES quartiles.

Dropout. Used both to describe an event (leaving school before graduating) and a status (an individual who was not in school and not a graduate at a defined point in time). The NELS:88 "cohort dropout rate" is based on a measurement of the enrollment status of 1988 8^{th}-graders 2 and 4 years later (in spring 1990 and spring 1992) and of 1990 sophomores 2 years later (in spring 1992). For a given point in time, a respondent is considered to be a dropout if he or she had not graduated from high school or attained an equivalency certificate and had not attended high school for 20 consecutive days (not counting excused absences). Transferring to another school is not regarded as a dropout event, nor is delayed graduation if a student was continuously enrolled but took an additional year to complete high school. A person who dropped out of school may have returned later and graduated. This person would be considered a "dropout" at the time he or she initially left school and a "stopout" at the time he or she returned to school.

4. SURVEY DESIGN

Target Population

Students enrolled in the 8^{th} grade in "regular" public and private schools located in the 50 states and the District of Columbia in the spring 1988 school term. The sample was freshened in both the first and second follow-ups to provide valid probability samples that would be nationally representative of 10^{th}-graders in spring 1990 and 12^{th}-graders in spring 1992. The NELS:88 project excludes the following types of schools: Bureau of Indian Education (BIE)[1] schools, special education schools for the handicapped, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas. The following students are also excluded: mentally handicapped students and students not proficient in English, for whom the NELS:88 tests would be unsuitable; and students having physical or emotional problems that would make participation in the survey unwise or unduly difficult. However, the Base-Year Ineligible Study (in the first follow-up) and the Followback Study of Excluded Students (in the second follow-up) sampled excluded students and added those no longer considered ineligible to the freshened sample of the first and second follow-ups, respectively.

Sample Design

NELS:88 was designed to follow a nationally representative longitudinal component of students who were in the 8^{th} grade in spring 1988. It also provides a nationally representative sample of schools offering 8^{th} grade in 1988. In addition, by freshening the student sample in the first and second follow-ups, NELS:88 provides nationally representative populations of 10^{th}-graders in 1990 and 12^{th}-graders in 1992. To meet the needs for cross-sectional, longitudinal, and cross-cohort analyses, NELS:88 involved complex research designs, including both longitudinal and cross-sectional sample designs.

Base-Year Survey. In the base year, students were selected using a two-stage stratified probability design, with schools as the first-stage units and students within schools as the second-stage units. From a national frame of about 39,000 schools with 8^{th} grades, a pool of 1,030 schools was selected through stratified sampling with probability of selection proportional to their estimated 8^{th}-grade enrollment; private schools were oversampled to ensure adequate representation. A pool of 1,030 replacement schools was selected by the same method to be used as substitutions for ineligible or

[1] These were referred to as Bureau of Indian Affairs (BIA) funded schools.

refusal schools in the initial pool. A total of 1,060 schools cooperated in the base year; of these, 1,060 schools (815 public and 237 private) contributed usable student data. The sampling frame for NELS:88 was the school database compiled by Quality Education Data, Inc., of Denver, Colorado, supplemented by racial/ethnic data obtained from the U.S. Office for Civil Rights and school district personnel.

Student sampling produced a random selection of 26,440 8th-graders in 1988; 24,600 participated in the base-year survey. Hispanic and Asian/Pacific Islander students were oversampled. Within each school, approximately 26 students were randomly selected (typically, 24 regularly sampled students and 2 oversampled Hispanic or Asian/Pacific Islander students). In schools with fewer than 24 8th-graders, all eligible students were selected. Potential sample members were considered ineligible and excluded from the survey if disabilities or language barriers were seen as obstacles to successful completion of the survey. The eligibility status of excluded members was reassessed in the first and second follow-ups. (See below.)

First Follow-up Survey. There were three basic objectives for the first follow-up sample design. First, the sample was to include approximately 21,500 students who were in the 8th-grade sample in 1988 (including base-year nonrespondents), distributed across 1,500 schools. Second, the sample was to constitute a valid probability sample of all students enrolled in the 10th grade in spring 1990. This entailed “freshening” the sample with students who were 10th-graders in 1990 but who were not in the 8th grade in spring 1988 or who were out of the country at the time of base-year sampling. The freshening procedure added 1,230 10th-graders; 1,040 of the students in this new group were found to be eligible and were retained after final subsampling for the first follow-up survey. Third, the first follow-up was to include a sample of students who had been deemed ineligible for base-year data collection due to physical, mental, or linguistic barriers to participation. The Base-Year Ineligible Study reassessed the eligibility of these students so that those able to take part in the survey could be added to the first follow-up student sample. Demographic and school enrollment information was also collected for all students excluded in the base year, regardless of their eligibility status for the first follow-up.

While schools covered in the NELS:88 base-year survey were representative of the national population of schools offering the 8th grade, the schools in the first follow-up were not representative of the national population of high schools offering the 10th grade. By 1990, the 1988 8th-graders had dispersed to many high schools, which did not constitute a national probability sample of high schools. To compensate for this limitation, the High School Effectiveness Study (HSES), which was designed to sustain analyses of school effectiveness issues, was conducted in conjunction with the first follow-up. From the pool of participating first follow-up schools, a probability subsample of 251 urban and suburban schools in the 30 largest Metropolitan Statistical Areas were designated as HSES schools. The NELS:88 core student sample was augmented to obtain a within-school representative student sample large enough to support school effects research. The student sample was increased in HSES schools by an average of 15 students to obtain within-school student cluster sizes of approximately 30 students.

Second Follow-up Survey. The second follow-up sample included all students and dropouts selected in the first follow-up. From within the schools attended by the sample members, 1,500 12th-grade schools were selected as sampled schools. Of these, the full complement of component activities occurred in 1,370 schools. For students attending schools other than these 1,370 schools, only the Student and Parent Questionnaires were administered. As in the first follow-up, the student sample was augmented through freshening to provide a representative sample of students enrolled in the 12th grade in spring 1992. Freshening added into the sample 243 eligible 12th-graders who were not in either the base-year or first follow-up sampling frames. Schools and students designated for the HSES in the first follow-up were followed up again—as part of both the main second follow-up survey and a separate HSES. The Followback Study of Excluded Students was a continuation of the first follow-up Base-Year Ineligible Study. In addition, two new components—the High School Transcript Study and the Course Offerings Component—were added to the second follow-up.

Third Follow-up Survey. The third follow-up student sample was created by dividing the second follow-up sample into 18 groups based on students’ response history, dropout status, eligibility status, school sector type, race, test scores, SES, and freshened status. Each sampling group was assigned an overall selection probability. Cases within a group were selected such that the overall group probability was met, but the probability of selection within the group was proportional to each sample member’s second follow-up design weight. Assigning selection probabilities in this way reduced the variability of the third follow-up raw weights and consequently increased the efficiency of the resulting sample from 40.1 to 44.0 percent.

Fourth Follow-up Survey. The fourth follow-up student sample was the same as the third follow-up student sample. Data collection for the NELS:88 fourth follow-up survey ended in September 2000, providing a final respondent population of approximately 12,100 individuals.

The Postsecondary Education Transcript Study, conducted as part of the fourth follow-up in 2000, followed those who reported having attended at least one postsecondary institution according to either the third follow-up survey in 1994 or the fourth follow-up survey in 2000. A total of approximately 9,600 fourth follow-up survey respondents (79 percent of the overall respondent population) reported postsecondary experience since high school. Approximately 21 percent of the NELS:88 respondent population did not participate in postsecondary education.

Within this sample of students, the transcript data collection further targeted students who attended only postsecondary institutions identified in the Integrated Postsecondary Education Data System (IPEDS) institutional data file, thus excluding postsecondary information collected from foreign institutions, non-degree-granting programs, and non-credit-granting institutions. Transcripts were requested from a total of 3,200 postsecondary institutions.

Data Collection and Processing

NELS:88 compiled data from five primary sources: students, parents, school administrators, teachers, and high school administrative records (transcripts, course offerings, and course enrollments). Data collection efforts for the base year through third follow-up extended from spring 1988 through summer 1994. Self-administered questionnaires, cognitive tests, and telephone or personal interviews were used to collect the data. The follow-up surveys involved extensive efforts to locate and collect data from sample members who were school dropouts, school transfers, or otherwise mobile individuals. Coding and editing conventions adhered as closely as possible to the procedures and standards previously established for NLS:72 and HS&B. The contractor National Opinion Research Center (NORC) at the University of Chicago was the prime contractor for the NELS:88 project from the base year through the third follow-up, but Research Triangle Institute conducted the fourth follow-up.

Reference dates. In the base-year survey, most questions referred to the student's experience up to the time of the survey administration in spring 1988. In the follow-ups, most questions referred to experiences that occurred between the previous survey and the current survey. For example, the second follow-up largely covered the period between 1990 (when the first follow-up was conducted) and 1992 (when the second follow-up was conducted).

Data collection. Prior to each survey, it was necessary to secure a commitment to participate in the study from the administrator of each sampled school. For public schools, the process began by contacting the Council of Chief State School Officers and the officer in each state. Once approval was gained at the state level, contact was made with district superintendents and then with school principals. For private schools, the National Catholic Educational Association and the National Association of Independent Schools were contacted for endorsement of the project, followed by contact of the school principals. The principal of each cooperating school designated a School Coordinator to serve as a liaison between contractor staff and selected respondents—students, parents, teachers, and the school administrator. The School Coordinator (most often a guidance counselor or senior teacher) handled all requests for data and materials, as well as all logistical arrangements for student-level data collection on the school premises. Coordinators were asked to identify students whose physical or learning disabilities or linguistic deficiencies would preclude participation in the survey and to classify all eligible students as White, Black, Hispanic, Asian/Pacific Islander, or "other" race.

For the base-year through second follow-up surveys, Student Questionnaires and test batteries were primarily administered in group sessions at the schools on a scheduled Survey Day. The sessions were monitored by contractor field staff, who also checked the questionnaires for missing data and attempted data retrieval while the students were in the classroom. Makeup sessions were scheduled for students who were unable to attend the first session. In the first and second follow-ups, off-campus sessions were used for dropouts and for sample members who were not enrolled in a first follow-up school on Survey Day. The School Administrator, Teacher, and Parent Questionnaires were self-administered. Contractor field staff followed up by telephone with individuals who had not returned their questionnaires by mail within a reasonable amount of time.

The first follow-up data collection required intensive tracing efforts to locate base-year sample members who, by 1990, were no longer in their 8th-grade schools but had dispersed to many high schools. Also, in order to derive a more precise dropout rate for the 1988 8th-grade cohort, a second data collection was undertaken 1 year later, in spring 1991. At this time, an attempt was made to administer questionnaires—by telephone or in person—to sample members who had missed data

collection at their school or who were no longer enrolled in school. The first follow-up also included the Base-Year Ineligible (BYI) Study, which surveyed a sample of students considered ineligible in the base year due to linguistic, mental, or physical deficiencies. The BYI Study sought to determine if eligibility status had changed for the excluded students so that newly eligible students could be added to the longitudinal sample. If an excluded student was now eligible, an abbreviated Student Questionnaire or a Dropout Questionnaire was administered, as appropriate. For those students who were still ineligible, their school enrollment status was ascertained and basic information about their sociodemographic characteristics was recorded.

Tracing efforts continued in the second and third follow-ups. In the second follow-up (conducted in 1992), previously excluded students were surveyed through the Followback Study of Excluded Students. The second follow-up also collected transcript, course offerings, and course enrollments from the high schools; reminder postcards were sent to principals who did not respond within a reasonable period. Data collection for the High School Effectiveness Study (HSES) was conducted concurrently with the collection for the second follow-up. Because of the overlap in school and student samples, survey instruments and procedures for the HSES were almost identical to those used in the NELS:88 second follow-up survey.

By 1994, when the third follow-up was conducted, most sample members had graduated from high school and it was no longer feasible to use group sessions to administer Student Questionnaires. Instead, the dominant form of data collection was one-on-one administration through computer-assisted telephone interviewing (CATI). In-person interviews were used for sample members who required intensive in-person locating or refusal conversion. Only the Student Questionnaire was administered in the third follow-up.

By 2000, when the fourth follow-up was conducted, most sample members who attended college and technical schools had completed their postsecondary education. Data collection for the fourth follow-up survey was conducted almost exclusively with computer-assisted interviewing, primarily by telephone (i.e., using CATI). However, in-person field interviews were also completed with this technology. Field interviewers used the same computer-assisted interview and online coding software as the study's telephone interviewers, but on a laptop computer-based platform (i.e., computer-assisted personal interviewing, or CAPI). Thus, all of the entry of interview data was accomplished by the NELS:88 fourth follow-up CATI-CAPI system.

High school transcripts were collected as part of the second follow-up. The groundwork for the collection of high school transcripts was laid in the spring and fall of 1991, during pre-data collection activities for the second follow-up. Principals were asked to provide any materials—such as course catalogs, student manuals or handbooks, course lists, and registration forms—that would aid transcript course coding. In August 1992, transcript survey materials were mailed to the principals of the NELS:88 and non-NELS:88 schools attended or most recently attended by sample members eligible for the survey. Two weeks after survey materials were mailed, nonresponding principals were prompted for the return of transcripts with a postcard reminder. Principals who did not return transcripts within 3 weeks of the postcard prompt were prompted over the telephone. Telephone prompting of nonresponding principals continued from October 1992 through February 1993. Field visits to schools requesting assistance in the preparation of transcripts were conducted in February and March.

The Postsecondary Education Transcript Study was carried out at the conclusion of CATI and CAPI data collection for the fourth follow-up survey. Data collection began in September 2000, and over the next 5 months project staff requested transcripts from postsecondary institutions that NELS:88 fourth follow-up respondents reported attending during either the NELS:88 third follow-up or NELS:88 fourth follow-up studies. Requests for transcripts were sent to the registrars or other contacts at the schools. Telephone follow-up with nonresponding institutions took place 2 weeks after transmission of the package. Data collection procedures were designed to follow, where possible, each institution's typical procedures for producing and distributing student transcripts. Returned transcripts and related school catalogs and bulletins were inventoried, transcript identification numbers affixed to each, and unique identifying information removed.

Processing. Data processing activities were quite similar for the base-year survey and the first and second follow-ups. An initial check of student documents for missing data was performed on-site by contractor staff so that data could be retrieved from the students before they left the classroom. Special attention was paid to a list of "critical items." Once the questionnaires and tests were received at the contractor, they were again reviewed for completeness, and a final disposition code was assigned to the case indicating which documents had been completed by the

sample member. Postsecondary institutions reported by the student were coded using the standard IPEDS codes. Data entry for both Student Questionnaires and cognitive tests was performed through optical scanning. New Student Supplements and Dropout Questionnaires were converted to machine-readable form using key-to-disk methods. All cognitive tests were photographed onto microfilm for archival storage. In the third follow-up, a CATI system captured the data at the time of the interview. The system evaluated the responses to completed questions and used the results to route the interviewer to the next appropriate question. The CATI program also applied the customary edits, described below under "Editing." At the conclusion of an interview, the completed case was deposited in the database ready for analysis. There was minimal post-data entry cleaning because the interviewing module itself conducted the majority of necessary edit checking and conversion functions.

Verbatim responses were collected in the third follow-up for a number of items, including occupation and major field of study. When respondents indicated their occupation, the CATI interviewers recorded the verbatim response. The system checked the response using a keyword search to match it to a subset of standard industry and occupation codes, and then presented the interviewer with a set of choices based on the keyword matches. The interviewer chose the option which most closely matched the information provided by the respondent, probing for additional information when necessary. Quality control was ensured by a reading and recoding, if necessary, of the verbatim responses by professional readers.

In the fourth follow-up, data were collected and edited almost exclusively with computer-assisted interviewing, primarily by telephone (i.e., using CATI).

For the High School Transcript Study, student- and course-level data were abstracted from transcripts. Transcript courses were coded using the course catalog for the school or district, in accordance with the Classification System of Secondary Courses, updated for the 1990 NAEP High School Transcript Study. When a school or district catalog was unavailable, courses were coded by title alone.

Information from the postsecondary education transcripts, including terms of attendance, fields of study, specific courses taken, and grades and credits earned, was coded and processed using a transcript control system developed specifically for this purpose. Specially trained research personnel then coded and tabulated these academic documents.

Editing. In the base-year through second follow-up surveys, detection of out-of-range codes was completed during scanning or data entry for all closed-ended questions. Machine editing was used to (1) resolve inconsistencies between filter and dependent questions; (2) supply appropriate missing data codes for questions left blank (e.g., legitimate skip, refusal); (3) detect illegal codes and convert them to missing data codes; and (4) investigate inconsistencies or contradictions. Frequencies and cross-tabulations for each variable were inspected before and after these steps to verify the accuracy and appropriateness of the machine editing. Items with unusually high nonresponse or multiple responses were further checked by verifying the responses on the questionnaire. A final editing step involved recoding Student Questionnaire responses for some items to the codes for the same items in earlier NELS:88 waves or in HS&B. Once this was done, codes that differed in the Dropout Questionnaire were recoded to coincide with the codes used for Student Questionnaire responses.

In the third and fourth follow-ups, machine editing was replaced by the interactive edit capabilities of the CATI system, which tested responses for valid ranges, data field size, data type (numeric or text), and consistency with other answers or data from previous rounds. If the system detected an inconsistency because of an interviewer's incorrect entry, or if the respondent simply realized that he or she had made a reporting error earlier in the interview, the interviewer could go back and change the earlier response. As the new response was entered, all of the edit checks performed at the first response were again performed. The system then worked its way forward through the questionnaire using the new value in all skip instructions, consistency checks, and the like until it reached the first unanswered question, and control was then returned to the interviewer. When problems were encountered, the system could suggest prompts for the interviewer to use to elicit a better or more complete answer.

Estimation Methods

Sample weighting is required so that NELS:88 data are representative of the full population. Imputation for missing nonresponses, however, has not yet been systematically provided for data analysis.

Weighting. Weighting is used in NELS:88 data analysis to accomplish a number of objectives, including (1) expanding counts from sample data to full population levels; (2) adjusting for differential selection probabilities (e.g., the oversampling of Asian and Hispanic students); (3) adjusting for differential response rates; and (4) improving representativeness by using auxiliary information. Multiple "final" (or

nonresponse-adjusted) weights have been provided for analyzing the different populations that NELS:88 data represent (i.e., base-year schools; 8^{th}-graders in 1988 and 2, 4, 6, and 12 years later; 1990 sophomores; 1992 seniors; and 2000 college graduates). Weights should be used together with the appropriate flag in order to analyze the sample for a particular targeted population.

Weights have not been constructed for all possible analytic purposes. In cases where no specific weight is available, existing weights may provide reasonable approximations. For instance, base-year parent and cognitive test completion rates were so high relative to Student Questionnaire completion that the student weight can be used for them with minimal bias.

NELS:88 weights were calculated in two steps: (1) unadjusted weights were calculated as the inverse of the probabilities of selection, taking into account all stages of the sample selection process; and (2) these initial weights were adjusted to compensate for nonresponse, typically carried out separately within multiple weighting cells. For detailed discussions of the calculation of weights for each wave, users are referred to the methodology reports for the individual surveys.

Scaling (Item Response Theory). Item Response Theory (IRT) was used to calibrate item parameters for all cognitive test items administered to students in NELS:88 tests. The tests conducted in each NELS:88 survey generated achievement measures in standardized scores.

Imputation. NELS:88 surveys have not involved large-scale imputation of missing data. Only a few variables have been imputed: student's sex, race/ethnicity, and school enrollment status. For example, when sex was missing in the data file, the information was looked for in earlier school rosters. If it was still unavailable after this review, sex was assumed from the sample member's name (if unambiguous). As a final resort, sex was randomly assigned.

5. DATA QUALITY AND COMPARABILITY

A number of studies have been conducted to address data quality issues relating to the NELS:88 project. During the course of data collection and processing, systematic efforts were made to monitor, assess, and maximize data quality. Subsequently, studies were conducted to evaluate the data quality in NELS:88 in comparison with that in earlier longitudinal surveys.

Sampling Error

Because the NELS:88 sample design involved stratification, disproportionate sampling of certain strata, and clustered (i.e., multistage) probability sampling, the calculation of exact standard errors (an indication of sampling error) for survey estimates can be difficult and expensive. For NELS:88, the Taylor series procedure has typically been used to calculate the standard errors.

Standard errors and design effects for about 30 key variables in each NELS:88 wave from the base year through the fourth follow-up were calculated using SUDAAN software. These can be used to approximate the standard errors if users do not have access to specialized software.

Design effects. A comparative study of design effects across NELS:88 waves and between NELS:88 and HS&B was done. When comparing NELS:88 base-year Student Questionnaire data to the results from HS&B—the 30 variables from the NELS:88 Student Questionnaire were selected to overlap as much as possible with those variables examined in HS&B—the design effects indicate that the NELS:88 sample was slightly more efficient than the HS&B sample. The smaller design effects in the NELS:88 base year may reflect its smaller cluster size (24 students plus, on average, two oversampled Hispanics and Asian from each NELS:88 school vs. the 36 sophomore and 36 senior selections from each HS&B school). The mean design effect for base-year students is 2.54.

In the comparative study of design effects across NELS:88 waves, the design effects in the subsequent follow-up studies were somewhat higher than those in the base year, a result of the subsampling procedures used in the follow-ups. The mean design effects for students and dropouts are 3.90 for the first follow-up, 3.70 for the second follow-up, 2.90 for the third follow-up, and 3.90 for the fourth follow-up. See the *NELS:88 Base Year Through Second Follow-up Final Methodology Report* (Ingels et al. 1998) and the *User's Manual: NELS:88 Base-Year to Fourth Follow-up: Student Component Data File* (Curtin et al. 2002).

Nonsampling Error

Coverage error. Exclusion and undercoverage of certain groups of schools and students in NELS:88 generated coverage error. In the base-year survey, for example, students who had linguistic, mental, or physical obstacles were excluded from the study.

Consequently, the national populations for such student groups were not fully covered by the sample.

To correct this coverage bias, the Base-Year Ineligible (BYI) Study collected eligibility information for 93.9 percent of the sample members excluded in the base-year survey. For those who were reclassified as eligible in the BYI Study, Student or Dropout Questionnaires were administered in person or over the telephone during the first follow-up. Cognitive tests were also administered to a small percentage of these students. For students who remained ineligible, school enrollment status and other key characteristics were obtained. The BYI Study permitted an evaluation of coverage bias in NELS:88 and a means of reducing undercoverage by identifying newly eligible students who could then be added into the sample to ensure cross-sectional representativeness. This effort also provided a basis for making corrected dropout estimates, taking into account both 1988-eligible and 1988-ineligible 8th-graders 2 years later. For further detail on the BYI Study, see *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years* (Ingels 1996).

Nonresponse error. Both unit nonresponse (nonparticipation in the survey by a sample member) and item nonresponse (missing value for a given questionnaire/test item) have been evaluated in NELS:88 data.

Unit nonresponse. In the NELS:88 base-year survey, the initial school response rate was 69 percent. This low rate prompted a follow-up survey to collect basic characteristics from a sample of the nonparticipating schools. These data were then compared to the same characteristics among the participating schools to assess the possible impact of response bias on the survey estimates. The school-level nonresponse bias was found to be small to the extent that schools could be characterized by size, control, organizational structure, student composition, and other factors. Bias at the school level was not assessed for the follow-up surveys because (1) sampling for the first and second follow-ups was student-driven (i.e., the schools were identified by following student sample members) and the third and fourth follow-ups did not involve schools; and (2) school cooperation rates were very high (up to 99 percent). Even if a school refused to cooperate, individual students were pursued outside of school (although school context data were not collected). The student response rates are shown in table 5 below.

Student-level nonresponse analysis was conducted with a focus on *panel nonresponse* since a priority of

Table 5. Unit-level and overall weighted response rates for selected NELS:88 student populations, by data collection wave

	Unit-level weighted response rate					
Population	Base-year school level	Base-year student level	1st follow-up	2nd follow-up	3rd follow-up	4th follow-up
Interviewed students	69.7 [1]	93.4	91.1	91.0	90.9	82.1
Tested students	69.7 [1]	96.5	94.1	76.6	†	†
Dropouts	69.7 [1]	†	91.0	88.0	†	†
Tested dropouts	69.7 [1]	†	48.6	41.7	†	†
	Overall weighted response rate					
	Base-year school level	Base-year student level	1st follow-up	2nd follow-up	3rd follow-up	4th follow-up
Interviewed students	69.7 [1]	65.1	63.5	63.4	63.4	57.2
Tested students	69.7 [1]	67.3	65.6	53.4	†	†
Dropouts	69.7 [1]	†	63.4	61.3	†	†
Tested dropouts	69.7 [1]	†	33.9	29.1	†	†

† Not applicable.
[1]Unweighted response rate.
SOURCE: Curtin, T.R., Ingels, S.J., Wu, S., and Heuer, R. (2002). *User's Manual: NELS:88 Base-Year to Fourth Follow-up: Student Component Data File* (NCES 2002-323). National Center for Education Statistics, U.S. Department of Education. Washington, DC. Spencer, B.D., Frankel, M.R., Ingel, S.J., Rasinski, K.A., and Tourangeau, R. (1990). *NELS:88 Base-Year Sample Design Report* (NCES 90-463). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

the NELS:88 project is to provide a basis for longitudinal analysis. Nonresponse was examined for the 8th-grade and 10th-grade cohorts. Any member of the 8th-grade cohort who did not complete a survey in three rounds (base year, first follow-up, and second follow-up) and any member in the 10th-grade cohort who did not complete a survey in the second and third rounds (first and second follow-ups) was considered a panel nonrespondent for that cohort. Panel nonresponse to cognitive tests in the two cohorts was defined the same way. The nonresponse rate was defined as the proportion of the selected students (excluding deceased students) who were nonrespondents in any round in which data were collected.

Nonresponse rates for both cohorts were calculated by school- and student-level variables that were assumed to be stable across survey waves (e.g., sex and race). These variables allowed comparisons between participants and nonparticipants even though the data for the latter were missing in some rounds. Estimates were made with both weighted and unweighted data. The weight used was the second follow-up raw panel weight (not available in the public-release dataset). About 18 percent of the 8th-grade cohort and 10 percent of the 10th-grade cohort were survey nonrespondents at one or more points in time. Approximately 43 percent of the 8th-grade cohort and 35 percent of the 10th-grade cohort did not complete one or more cognitive tests in their rounds of testing.

Nonresponse bias was calculated as the difference in the estimates between the respondents and all selected students. On the whole, the analysis revealed only small discrepancies between the two cohorts. Bias estimates were higher, however, for the 8th-grade cohort than for the 10th-grade cohort because of the 8th-grade cohort's more stringent definition of participation. The discrepancies between cognitive test completers and noncompleters were larger than between survey participants and nonparticipants; this pattern held for both cohorts. In brief, the magnitude of the bias was generally small—few percentage estimates were off by as much as 2 percent in the 8th-grade cohort and 1 percent in the 10th-grade cohort. Such bias reflects the raw weight. The nonresponse-adjusted weight should correct for differences by race and sex to produce correct population estimates for each subgroup.

Further analysis was done using several other student and school variables. The results showed rather similar patterns of bias. When compared with estimates from HS&B, the student nonresponse bias estimates in NELS:88 were consistently lower. However, the two studies seem to share certain common patterns of nonresponse. For example, both studies generated comparatively higher nonresponse rates among students enrolled in schools in the West, Black students, students in vocational or technical programs, students in the lowest test quartile, and dropouts.

Item nonresponse. Item nonresponse was examined in base-year through second follow-up data obtained from surveys of students, parents, and teachers. Differences emerged among student subgroups in the level of nonresponse to a wide range of items—from language background, family composition, and parents' education to perception of school safety. Nonresponse was often two to five times as great for one subgroup as for the other subgroups. High item nonresponse rates were associated with such attributes as not living with parents, having low SES, being male, having poor reading skills, and being enrolled in a public school. Compared with parent nonresponse to items about college choice and occupational expectations, student nonresponse rates were generally lower. For items about student's language proficiency, classroom practices, and student's high school track, students had consistently lower nonresponse rates than their teachers did. See the *NELS:88 Survey Item Evaluation Report* (McLaughlin, Cohen, and Lee 1997) for further detail.

Measurement error. NCES has conducted studies to evaluate measurement error in (1) student data (compared to parent and teacher data); and (2) student cognitive test data.

Parent-student convergence and teacher-student convergence. A study of measurement error in data from the base-year through second follow-up surveys focused on the convergence of responses by parents and students and by teachers and students. (See the *NELS:88 Survey Item Evaluation Report* [McLaughlin, Cohen, and Lee 1997].) Response convergence (or discrepancy) across respondent groups can be interpreted as an indication of measurement reliability, validity, and communality, although the data are often not sufficient to determine which response is more accurate.

The student and parent components of this study covered such variables as number of siblings, the student's work experience, language background, parents' education, parent-student discussion of issues, perceptions about school, and college and occupation expectations. Parent-student convergence varied from very high to very low, depending on the item. For example, convergence was high for number of siblings, regardless of student-level characteristics such as SES, sex, reading scores, public versus private school

enrollment, and whether or not living with parents. In contrast, parent-student convergence was low for items related to the student's work experience; there was also more variation across student subgroups for these items. In general, convergence tended to be high for objective items, for items worded similarly, and for nonsensitive items.

Teacher-student convergence was examined through variables about student's English proficiency, classroom practices, and student's high school track. Again, convergence was found to vary considerably across data items and student subgroups. Convergence was high for student's native language but low for student's English proficiency. Across student subgroups, there was a greater range in correlations for English proficiency than for native language. Teachers and students differed quite dramatically on items about classroom practices.

Cognitive test data. In-depth studies of measurement error issues related to cognitive tests administered in the base-year through second follow-up surveys are also available. See the *Psychometric Report for the NELS:88 Base Year Test Battery* (Rock and Pollack 1991) and the *Psychometric Report for the NELS:88 Base Year Through Second Follow-up* (Rock and Pollack 1995).

The first study (Rock and Pollack 1991) addressed issues related to test speediness (the limited testing time in relation to the outcome), reliability, item statistics, performance by racial/ethnic and gender groups, and IRT parameters for the battery. The results indicate that the test battery either met or exceeded all of its psychometric objectives. Specifically, the study reported: (1) while the allotted testing time was only 1½ hours, quite acceptable reliability was obtained for the tests on reading comprehension, mathematics, history/citizenship/ geography, and, to a somewhat lesser extent, science; (2) the internal consistency reliability was sufficiently high to justify the use of IRT scoring and, thus, provide the framework for constructing 10th- and 12th-grade forms that would be adaptive to the ability levels of the students; (3) there was no consistent evidence of differential item functioning (item bias) for gender or racial/ethnic groups; (4) factor analysis results supported the discriminant validity of the four tested content areas; convergent validity was also indicated by salient loadings of testlets composed of "marker items" on their hypothesized factors; and (5) in addition to providing the usual normative scores in all four tested areas, behaviorally anchored proficiency scores were provided in both the reading and math areas.

The second study (Rock and Pollack 1995) focused on issues relating to the measurement of gain scores. Special procedures were designed into the test battery design and administration to minimize the floor and ceiling effects that typically distort gain scores. The battery used a two-stage multilevel procedure that attempted to tailor the difficulty of the test items to the performance level of a particular student. Thus, students who performed very well on their 8th-grade mathematics test received a relatively more difficult form in 10th grade than students who had not performed well on their 8th-grade test. There were three forms of varying difficulty in mathematics and two in reading in both grades 10 and 12. Since 10th- and 12th-graders were taking forms that were more appropriate for their level of ability and achievement, measurement accuracy was enhanced and floor and ceiling effects could be minimized. The remaining two content areas—science and history/citizenship/geography—were only designed to be grade-level adaptive (i.e., a different form for each grade but not multiple forms varying in difficulty within grade).

To maximize the gain from using an adaptive procedure, special vertical scaling procedures were used that allow for Bayesian priors on subpopulations for both item parameters and scale scores. In comparing more traditional non-Bayesian approaches to scaling longitudinal measures with the Bayesian approach, it was found that the multilevel approach did increase the accuracy of the measurement. Furthermore, when used in combination with the Bayesian item parameter estimation, the multilevel approach reduced floor and ceiling effects when compared to the more traditional IRT approaches.

Data Comparability

NELS:88 is designed to facilitate both longitudinal and trend analyses. Longitudinal analysis calls for data compatibility across survey waves whereas trend analysis requires data compatibility with other longitudinal surveys. Data compatibility issues may relate to survey instruments, sample design, and data collection methods.

Comparability within NELS:88 across survey waves. A large number of variables are common across survey waves. (See the *NELS:88 Second Follow-up Student Component Data File User's Manual* [Ingels et al. 1994] for a listing of common Student Questionnaire variables in the base year, first follow-up, and second follow-up.) However, compatibility of NELS:88 data across waves can still be an issue because of subtle differences in question wording, sample differences (e.g., with or without dropouts and freshening students, sample attrition, nonresponse), and data collection

methods (e.g., on-campus group session, off-campus individual survey, telephone interview).

One NCES study compared 112 pairs of variables repeated from the base year to the first and second follow-up surveys. (See the *NELS:88 Survey Item Evaluation Report* [McLaughlin, Cohen, and Lee 1997].) These variables cover student family, attitudes, education plans, and perceptions about schools. The results suggest that the interpretations of NELS:88 items depend on the age level at which they were administered. Data convergence tended to be higher for pairs of first and second follow-up measures than for pairs of base-year and second follow-up measures. Some measures were more stable than others. Students responded nearly identically to the base-year and second follow-up questions about whether English was their native language. Their responses across survey waves were also fairly stable as to whether their curriculum was intended to prepare them for college, whether they planned to go to college, and their religiosity. It should be noted that cross-wave discrepancies may reflect a change in actual student behavior rather than a change in response for a status quo situation.

Comparability within NELS:88 across respondent groups. While different questionnaires were used to collect data from different respondent groups (students, parents, teachers, school administrators), there are overlapping items among these instruments. One study examined the extent to which the identical or similar items in different questionnaires generated compatible information. It found considerable discrepancies between students and parents, and even greater discrepancies between students and teachers, in their responses to selected groups of overlapping variables. (See “Measurement error” above.)

Comparability with NLS:72, HS&B, and ELS:2002. NELS:88 surveys contain many items that are also covered in NLS:72, HS&B, and ELS:2002—a feature that enables trend analyses of various designs. (See the *NELS:88 Second Follow-up Student Component Data File User’s Manual* [Ingels et al. 1994] for a cross-walk of common variables and a discussion of trend analyses.) To examine data compatibility across the four studies, one should consider their sample designs and data contents, including questionnaires, cognitive tests, and transcript records.

Sample designs for the four studies are similar. In each base year, students were selected through a two-stage stratified probability sample, with schools as the first-stage units and students within schools as the second-stage units. In NLS:72, all baseline sample members were spring term 1972 high school seniors. In HS&B, all members of the student sample were spring term 1980 sophomores or seniors. In ELS:2002, the base-year sample students were 10th-graders. Because NELS:88 base-year sample members were 8th-graders in 1988, its follow-ups encompass students (both in the modal grade progression sequence and out of sequence) and dropouts. Sample freshening was used in NELS:88 to provide cross-sectional nationally representative samples. Despite similarities, however, the sample designs of the four studies differ in three major ways: (1) the NELS:88 first and second follow-ups had relatively variable, small, and unrepresentative within-school student samples, compared to the relatively uniform, large, and representative within-school student samples in NLS:72 and HS&B; (2) unlike the two earlier studies, NELS:88 did not provide a nationally representative school sample in its follow-ups; and (3) there were differences in school and subgroup sampling and oversampling strategies in the four studies. These sample differences imply differences in the respondent populations covered by the four studies.

Questionnaire overlap is apparent among the four studies; nevertheless, caution is required when making trend comparisons. Some items were repeated in identical form across the studies; others appear to be essentially similar but have small differences in wording or response categories.

IRT scaling was used in the four studies to put math, vocabulary, and reading *test scores* on the same scale for 1972, 1980, 1982, and 2002 seniors. Additionally, there were common items in the HS&B and NELS:88 math tests that provide a basis for equating 1980–1990 and 1982–1992 math results, and common items in the NELS:88 and ELS:2002 reading and math tests that provide the link to obtain the ELS:2002 student ability estimates on the NELS:88 ability scale. In general, however, the tests in the four studies differed in many ways. Although group differences by standard deviation units may profitably be examined, caution should be exercised in drawing time-lag comparisons for cognitive test data.

Transcript studies in NELS:88, HS&B, ELS:2002, and NAEP were designed to support cross-cohort comparisons. The ELS:2002, NAEP, and NELS:88 studies, however, provide summary data in Carnegie units, whereas HS&B provides course totals. Note too that course offerings were only collected from schools that were part of the High School Effectiveness Study in the NELS:88 second follow-up, whereas course offerings were collected from *all* schools in HS&B (see chapter 7), and course offerings were collected from all base-year schools and the last school attended by

sample members who transferred out of their base-year school in ELS:2002 (see chapter 9).

Other factors should also be considered in assessing data compatibility. Differences in mode and time of survey administration across the cohorts may affect compatibility. NELS:88 seniors were generally surveyed earlier in the school year than were NLS:72 seniors. NLS:72 survey forms were administered by school personnel while HS&B and NELS:88 survey forms were administered primarily by contractor staff. There were also differences in questionnaire formats; the later tests had improved mapping and different answer sheets.

6. CONTACT INFORMATION

For content information on the NELS:88 project, contact:

Jeffrey Owings
Phone: (202) 502-7423
E-mail: jeffrey.owings@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Alt, M.N., and Bradby, D. (1999). *Procedures Guide for Transcript Studies* (NCES Working Paper 1999-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Haggerty, C., Dugoni, B., Reed, L., Cederlund, A., and Taylor, J. (1996). *National Education Longitudinal Study (NELS:88/94) Methodology Report* (NCES 96-174). National Center for Education Statistics. U.S. Department of Education. Washington, DC.

Horn, L., and West, J. (1992). *A Profile of Parents of Eighth Graders* (NCES 92-488). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S.J., Dowd, K.L., Stipe, J.L., Baldridge, J.D., Bartot, V.H., and Frankel, M.R. (1995). *User's Manual: NELS:88 Second Follow-Up: Dropout Component Data File* (NCES 93-375). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S.J., Scott, L.A., Taylor, J.R., Owings, J., and Quinn, P. (1998). *NELS:88 Base Year Through Second Follow-Up: Final Methodology Report* (NCES Working Paper 98-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Uses of Data

Curtin, T.R., Ingels, S.J., Wu, S., and Heuer, R. (2002). *User's Manual: NELS:88 Base-Year to Fourth Follow-up: Student Component Data File* (NCES 2002-323). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S., and Baldridge, J. (1995). *National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors* (NCES Working Paper 95-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S., and Dowd, K. (1995). *National Education Longitudinal Study of 1988: Conducting Trend Analyses: HS&B and NELS:88 Sophomore Cohort Dropouts* (NCES Working Paper 95-07). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S., and Taylor, J. (1995). *National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data* (NCES Working Paper 95-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Survey Design

Bradby, D., and Hoachlander, G. (1999). *1998 Revision of the Secondary School Taxonomy* (NCES Working Paper 1999-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Quinn, P. (1995). *NELS:88 Second Follow-up Research and Development Working Papers* (NCES 94-251). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S. (1996). *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years* (NCES 96-723). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Rock, D.A., and Pollack, J.M. (1991). *Psychometric Report for the NELS:88 Base Year Test Battery* (NCES 91-468). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A., and Tourangeau, R.E. (1990). *NELS:88 Base Year Sample Design Report* (NCES 90-463). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Wu, S., Levy, P.S., Wang, J., Chromy, J., and Li, L. (2003). *Report for Computation of Balanced Repeated Replicate (BRR) Weights for the Third (NELS88:1994) and Fourth (NELS88:2000) Follow-up Surveys* (NCES Working Paper 2003-18). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Bokossa, M.C., and Huang, G.G. (2001). *Imputation of Test Scores in the National Education Longitudinal Study of 1988 (NELS:88)* (NCES Working Paper 2001-16). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Broene, P., and Rust, K. (2000). *Strengths and Limitations for Using SUDAAN, Stata, and WesVarPC for Computing Variances From NCES Data Sets* (NCES Working Paper 2000-03). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Burkam, D.T. (2003). *English Coursetaking and the NELS:88 Transcript Data* (NCES Working Paper 2003-02). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Burkam, D.T., and Lee, V. (2003). *Mathematics, Foreign Language, and Science Coursetaking and the NELS:88 Transcript Data* (NCES Working Paper 2003-01). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kaufman, R., and Rasinski, K.A. (1991). *Quality of Responses of Eighth-Grade Students in NELS:88* (NCES 90-487). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

McLaughlin, D.H., Cohen, J., and Lee. R. (1997). *NELS:88 Survey Item Evaluation Report* (NCES 97-052). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Pollack, J.M., and Rock, D.A. (1997). *Constructed Response Tests in the NELS:88 High School Effectiveness Study* (NCES 97-804). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Rock, D.A., and Pollack, J.M. (1995). *Psychometric Report for the NELS:88 Base Year Through Second Follow-up* (NCES 95-382). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Salvucci, S., Walter, E., Colney, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Spencer, B.D., Frankel, M. R., Ingel, S.J., Rasinski, K.A., and Tourangeau, R. (1990). *NELS: 88 Base-Year Sample Design Report* (NCES 90-463). National Center for Education Statistics, U.S. Department of Education.Washington, DC.

Chapter 9: Education Longitudinal Study of 2002 (ELS:2002)

1. OVERVIEW

LONGITUDINAL SAMPLE SURVEY OF THE 10TH-GRADE CLASS OF 2002; BASE-YEAR SURVEY, FIRST FOLLOW-UP IN 2004, AND SECOND FOLLOW-UP IN 2006

ELS:2002 collects data from:

- Students and dropouts
- School administrators
- Teachers
- Library media staff
- School facility checklist
- Parents
- High school transcripts

The Education Longitudinal Study of 2002 (ELS:2002) represents a major longitudinal effort designed to provide trend data about critical transitions experienced by students as they proceed through high school and into postsecondary education or their careers. The 2002 sophomore cohort is being followed, initially at 2-year intervals, to collect policy-relevant data about educational processes and outcomes, especially as such data pertain to student learning, predictors of dropping out, and high school effects on students' access to, and success in, postsecondary education and the workforce.

In the spring term of 2002 (the base year of the study), high school sophomores were surveyed and assessed in a national sample of high schools with 10th grades. Their parents, teachers, principals, and librarians were surveyed as well.

In the first of the follow-ups, base-year students who remained in their base-year schools were resurveyed and tested (in mathematics) 2 years later, along with a freshening sample that makes the study representative of spring 2004 high school seniors nationwide. Students who had transferred to a different school, switched to a homeschool environment, graduated early, or dropped out were administered a questionnaire. In the second follow-up in 2006, information was collected through a single electronic questionnaire about colleges applied to and aid offers received, enrollment in postsecondary education, employment and earnings, and living situation, including family formation. The third follow-up is planned for 2012, so that later outcomes, such as their persistence and attainment in higher education, or their transition into the labor market, can be understood in terms of their earlier aspirations, achievement, and high school experiences.

Purpose

ELS:2002 is designed to monitor the transition of a national sample of young people as they progress from 10th grade through high school and on to postsecondary education and/or the world of work.

Components

ELS:2002 has two distinctive features. First, it is a longitudinal study in which the same units are surveyed repeatedly over time. Individual students will be followed for more than 10 years; the base-year schools were surveyed two times, once in 2002 and again in 2006. Second, in the high school years, it is an integrated multilevel study that involves multiple respondent populations. The respondents include students, their parents, their teachers, their librarians, and their schools.

Base-Year Survey. The base-year (2002) data collection instruments for ELS:2002 consisted of five separate questionnaires (student, parent, teacher, school administrator, and library media center), two achievement tests (assessments in reading and mathematics), and a school observation form (facilities checklist).

Student Questionnaire. The student questionnaire gathered information about the student's background, school experiences and activities, plans and goals for the future, employment and out-of-school experiences, language background, and psychological orientation toward learning. The student questionnaire was divided into seven sections: (1) locating information, (2) school experiences and activities, (3) plans for the future, (4) non-English language use, (5) money and work, (6) family, and (7) beliefs and opinions about self. Assessments in reading and mathematics were given at the same time. The baseline scores for the assessments can serve as a covariate or control variable for later analyses. Mathematics achievement was reassessed 2 years later, so that achievement gain over the last 2 years of high school could be measured and related to school processes and mathematics coursetaking.

Parent Questionnaire. One parent of each participating sophomore was asked to respond to a parent survey. The parent questionnaire was designed to gauge parents' aspirations for their child and to collect information about the home background and home education support system, the child's educational history prior to 10th grade, and parents' interactions with and opinions about the student's school.

Teacher Questionnaire. For each student enrolled in English or mathematics, a teacher was also selected to participate in a teacher survey. The teacher questionnaire was designed to illuminate questions on the quality, equality, and diversity of educational opportunity by obtaining information in two content areas: the teacher's evaluations of the student and information about the teacher's background and activities.

School Administrator Questionnaire. The school administrator questionnaire collected information on school characteristics, student characteristics, teaching staff characteristics, school policies and programs, technology, and school governance and climate. The school administrator data can be used contextually, as an extension of the student data, when the student is the fundamental unit of analysis. At the same time, the data from the school administrator questionnaire are nationally representative and can be used to generalize to the nation's regular high schools with sophomores in the 2001–02 school year.

Library Media Center Questionnaire. For the school library media center component, the school librarian, media center director, or school administrator supplied information about library media center size, organization, and staffing; technology resources and electronic services; the extent of library and media holdings, including both collections and expenditures; and levels of facility utilization, including scheduling for use by students and teachers. Finally, the questionnaire supplied information about the library media center's use in supporting the school's curriculum; that is, how library media center staff collaborate with and support teachers to help them plan and deliver instruction. Information in the library media center questionnaire can be used as contextual data with the student as the unit of analysis or to generalize to libraries within all regular high schools with 10th grades in the United States in the 2001–02 school year.

School Facilities Checklist. The facilities component comprised a checklist to be completed by the survey administrator. The survey administrator was asked to observe a number of conditions at the school, including the condition of the hallways, main entrance, lavatories, classrooms, parking lots, and surrounding neighborhood. Of special interest were indicators of security (metal detectors, fire alarms, exterior lights, fencing, security cameras, etc.) and maintenance and order (trash, graffiti, clean walls and floors, noise level, degree of loitering, etc.). Information gathered in the facilities checklist can be used as contextual data with the student as the unit of analysis, or data can be used at the school level to generalize to all regular high schools with 10th grades in the United States in the 2001–02 school year.

First Follow-up Survey. The first follow-up (2004) survey comprised seven questionnaires and an achievement test in mathematics. The questionnaires included a student questionnaire, a transfer student questionnaire, a new participant supplement questionnaire (NPSQ) (repeating selected questions from the base year), a homeschool student questionnaire, an early graduate questionnaire, a dropout (not currently in school) questionnaire, and a school administrator questionnaire.

Student questionnaire. The student questionnaire was administered to sophomore cohort members who had remained in their base-year school as well as to a freshening sample of 12th-graders in the same schools. Students who completed the student questionnaire also were normally eligible for the first follow-up mathematics assessment. Some students were administered an abbreviated version of the questionnaire. The full questionnaire comprised eight content modules: (1) contact information in support of the longitudinal design; (2) the student's school experiences and activities, including information about extracurricular participation, computer use in English and math, the transition process from the sophomore

year to upper-level secondary school, and the relationship of curricular programs and coursetaking to educational achievement and persistence; (3) time usage on homework, TV viewing, video and computer games, computers, nonschool reading, library utilization, and other activities; (4) plans and expectations for the future, including students' educational and life goals and values; (5) education after high school; (6) plans for work after high school; (7) work status and history; and (8) community, family, and friends.

Transfer student questionnaire. Sophomore cohort members who had transferred out of their base-year school to a new school received the transfer student questionnaire. Transfer students were asked a subset of items from the student questionnaire covering the following topics: school experiences and activities; time use; plans and expectations for the future; education after high school; work after high school; and community, family, and friends. In addition, transfer students were asked when they transferred and their reasons for doing so. Transfer students did not complete a cognitive test, but their test scores have been imputed.

New participant supplement questionnaire (NPSQ). Any student new to the study at any of the core (base-year) schools was administered the NPSQ. The NPSQ gathered information (that had been collected for other students in the base year) on new participants' demographic characteristics, parental education and occupation, and language use. In addition, a subset of items included in the student questionnaire was also posed to new participants. These items (which are identical in content to those in the abbreviated student questionnaire) relate to topics such as school experiences and activities; time use; plans and expectations for the future; education and work after high school; and work, community, family, and friendship experiences. In contrast, the New Participant Supplement (NPS) gathered the key base-year variables that also were included in the NPSQ.

Homeschool student questionnaire. ELS:2002 does not provide a representative sample of homeschooled high school students. (In the base year, all study sophomores were selected from regular U.S. high schools.) Instead, homeschooled students in ELS:2002 generalize only to sophomores in regular high schools in the spring term of 2002 who were in a homeschool situation 2 years later. Homeschooled students were asked about their schooling activities and status, including their grade, coursework completed in science and math, and steps taken toward college; how they spend their time; their plans and expectations for the future, including education and work after high school; work experiences; and community, family, and friends.

Early graduate questionnaire. Early graduates were defined as sophomore cohort members who had graduated from high school or received a General Educational Development (GED) credential on or before March 15, 2004. Early graduates completed only a subset of the items in the student questionnaire, complemented by additional items pertaining to their situation. More specifically, early graduates were asked with whom they consulted when deciding to graduate early, the basis for that decision, and the means by which they did so. They also provided a history of their work and educational experiences since leaving high school.

Dropout questionnaire. Dropouts were defined as sophomore cohort members who were out of school in the spring term of 2004, who had not received a high school diploma or GED credential, and who had missed 4 or more consecutive weeks to a cause other than accident or illness. There was considerable overlap between the student and dropout questionnaires; both collected locating information for longitudinal follow-up and included items on school experiences and activities, time use, plans and expectations for the future, and the type and amount of work in which dropouts were engaged. The dropout questionnaire gathered information about students' work status and history, volunteer work or community college experience, and the educational behavior of friends. In the area of school experiences and activities, dropouts were asked questions about the school they last attended and their participation in alternative education programs. In addition, they were asked to supply their specific reasons for leaving school prior to graduation. They were asked as well about plans to get a GED or return to high school.

School administrator questionnaire content and content linkages. The school administrator questionnaire collected information on the school in four areas: school characteristics, structure, and policies; student characteristics and programs; teacher and library staff characteristics; and principal reports on the school environment. It should be noted that school-level data are not nationally representative of American high schools in 2004, since the first follow-up sample did not factor in "births" of new schools and "deaths" of existing schools between 2002 and 2004. First follow-up school data, however, do provide a statistical portrait of a nationally representative sample of American high schools with 10th grades in 2002 (2 years later).

Second Follow-up Survey. The second follow-up (2006) survey was a single electronic questionnaire administered in three modalities—a web-enabled self-administration, computer-assisted telephone interviewing (CATI), and computer-assisted personal interviewing (CAPI). (Both CATI and CAPI are interviewer-administered modalities.) The questionnaire covered the transition from high school to postsecondary education, and included items on college access and choice. Items were drawn from a number of studies, including the Baccalaureate and Beyond Longitudinal Study (B&B, see chapter 16), Beginning Postsecondary Students (BPS, see chapter 15) Longitudinal Study, High School and Beyond (HS&B) Longitudinal Study (see chapter 7), National Education Longitudinal Study of 1988 (NELS:88, see chapter 8), and National Postsecondary Student Aid Study (NPSAS, see chapter 14). The interview was organized into four substantive sections: *High School*, *Postsecondary Education, Employment,* and *Community*. The interview concluded with a *Locating* section.

The first section, *High School*, collected retrospective information about high school completion. Respondents were classified as spring-term 2004 12th-graders, spring-term 2004 dropouts, neither, or both (for a small set). The majority of respondents skipped this section entirely because their high school completion date and the type of high school credential they earned were preloaded into the instrument at the start of data collection.

The *Postsecondary Education* section of the interview, the point of entry for most respondents, focused on education *after* high school. Questions pertained to the application process, admissions, financial aid offers, institutions attended, experiences at these institutions, and educational expectations. Complete month-by-month enrollment histories for all postsecondary institutions attended after high school were collected in this section. These enrollment histories (in conjunction with the date of high school completion or exit, as preloaded or reported in the *High School* section of the interview) were used to classify respondents into one of six mutually exclusive categories: standard enrollees, delayers, leavers, delayer-leavers, nonenrollees, and high school students. The questions administered to each respondent depended on his/her category. These categories were used for the *Employment* and *Community* sections as well. For more details, see the *Education Longitudinal Study of 2002: Base-Year to Second Follow-up Data File Documentation* (Ingels et al. 2007).

There were five topics in the *Employment* section. The questions for the first topic referred to the first job after high school. The second set of questions focused on employment at the time of the interview. The next set focused on jobs held by postsecondary students during the 2004–05 and 2005–06 academic years. Respondents were also questioned about months of unemployment (if a gap existed between high school and their first job, their first job and their current job, and/or their first job and the date of the interview, if they were not currently working). Lastly, the questions for the fifth topic focused on income, finances, and occupational expectations at age 30.

The final substantive section of the interview, *Community*, covered topics related to family formation, living arrangements, community involvement (including military service), and experiences that may influence the life course. With one minor exception, all questions pertained to all respondent types.

The interview concluded with the *Locating* section, which collected information that will be used to contact the respondents in the next round of the study.

High School Transcript Study. Transcripts were collected from sample members in late 2004 and early 2005, about 6 months to 1 year after most students had graduated from high school. Transcripts were collected from the students' base-year school. However, if it was learned during the first follow-up data collection that they had transferred, transcripts were collected from two schools: the base-year school and the last known school of attendance. For students who were added to the study during their senior year (known as "freshened" students), transcripts were only collected from their senior-year school. Transcripts were collected for regular graduates, as well as dropouts, early graduates, and students who were homeschooled after their sophomore year. For more information, see Chapter 29, High School Transcript (HST) Studies.

The ELS:2002 high school transcript data collection sought key pieces of information about coursetaking from students' official high school records (e.g., courses taken while attending secondary school, credits earned, year and term a specific course was taken, and final grades). When available, other information, such as dates enrolled, reason for leaving school, and standardized test scores, was collected. All information was transcribed and can be linked back to the students' questionnaire or assessment data. Because of the size and complexity of the file and the reporting variation by school, additional variables were constructed from the raw transcript file to facilitate analyses. These variables include standardized grade point averages

(GPAs), academic pipeline measures, and total credits earned by subject area. The construction of many of the transcript variables is based on Carnegie units. A Carnegie unit is equal to a course taken every day, one period per day, for a full school year.

Third Follow-up Survey. The third follow-up is planned for 2012. By this time, most of those who attended college will have graduated and entered the labor market. The third follow-up will collect data on the post-high school educational experiences of all sample members (such as their postsecondary persistence, attainment), their history of employment, family formation, community service, and other areas. Postsecondary transcripts will be obtained as well.

Periodicity

The base-year survey was conducted in the spring of 2002. The first follow-up was done in 2004, as was the high school transcript component. A post-high school follow-up was done in 2006. The third follow-up is planned for 2012; in this final follow-up, college transcripts will be obtained.

2. USES OF DATA

Using the multilevel and longitudinal information from the base year (2002) and first follow-up (2004) of ELS:2002 will help researchers and policymakers explore and better understand such issues as the importance of home background and parental aspirations for a child's success; the influence of different curriculum paths and special programs; the effectiveness of different high schools; and whether a school's effectiveness varies with its size, organization, climate or ethos, curriculum, academic press, or other characteristics. These data will facilitate an understanding of the impact of various instructional methods and curriculum content and exposure in bringing about educational growth and achievement.

After the high school years, ELS:2002 will continue to follow its sample of students into postsecondary education and/or the labor market. For students who continue on to higher education, data collected from the second follow-up and the third follow-up (which is planned for 2012) will help researchers measure the effects of these students' high school careers on subsequent access to postsecondary institutions; their choices of institutions and programs; and, as time goes on, their postsecondary persistence, attainment, and eventual entry into the labor force and adult roles. For students who go directly into the workforce (whether as dropouts or high school graduates), ELS:2002 will be able to determine how well high schools have prepared these students for the labor market and how they fare within it.

Apart from helping to describe the status of high school students and their schools, the second and third follow-up data will provide information to help address a number of key policy and research questions. The study is intended to produce a comprehensive dataset for the development and evaluation of education policy at all government levels. Part of its aim is to inform decisionmakers, educational practitioners, and parents about the changes in the operation of the education system over time and the effects of various elements of the system on the lives of the individuals who pass through it. Issues that can be addressed with data collected in the high school years include the following:

- students' academic growth in mathematics;
- the process of dropping out of high school—determinants and consequences;
- the role of family background and the home education support system in fostering students' educational success;
- the features of effective schools;
- the impact of coursetaking choices on success in the high school years (and thereafter);
- the equitable distribution of educational opportunities as registered in the distinctive school experiences and performance of students from various subgroups; and
- steps taken to facilitate the transition from high school to postsecondary education or the world of work.

After ELS:2002 students have completed high school, a new set of issues can be examined using data from the second and third follow-ups. These issues include

- the later educational and labor market activities of high school dropouts;
- the transition of students who do not go directly on to postsecondary education or the world of work;
- access to, and choice of, undergraduate and graduate education institutions;

- persistence in attaining postsecondary educational goals;
- rate of progress through the postsecondary curriculum;
- degree attainment;
- barriers to persistence and attainment;
- entry of new postsecondary graduates into the workforce;
- social and economic rate of return on education to both the individual and society; and
- adult roles, such as family formation and civic participation.

3. KEY CONCEPTS

Cognitive Test Battery.
The test questions were selected from previous assessments: NELS:88, the National Assessment of Educational Progress (NAEP, see chapter 18), and Program for International Student Assessment (PISA, see chapter 22). Most, but not all, were multiple choice items. Test specifications for ELS:2002 were adapted from frameworks used for NELS:88. Math tests contained items in arithmetic, algebra, geometry, data/probability, and advanced topics were divided into process categories of skill/knowledge, understanding/ comprehension, and problem solving. Through inclusion of items from the PISA, the ELS:2002 math tests placed a somewhat greater emphasis on practical applications and problem solving than did the NELS:88 test forms. Reading tests consisted of reading passages of one paragraph to one page in length, followed by three to six questions based on each passage. The reading passages included literary material as well as topics in the natural and social sciences. Several passages required interpretation of graphs. Questions were categorized as reproduction of detail, comprehension, or inference/evaluation.

Cohort. A cohort is a group of individuals who have a statistical factor in common; for example, year of birth, grade in school, or year of high school graduation. ELS:2002 is a sophomore-grade cohort based on the spring term of the 2001–02 school year. It also contains, however, a nationally representative sample of high school seniors in the spring term of the 2003–04 school year.

Socioeconomic Status (SES). A composite variable is constructed through the combination of two or more variables—socioeconomic status, for example, combines mother's education, father's education, mother's occupation, father's occupation, and family income or an income proxy (household items) or it is calculated through the application of a mathematical function or transformation to a variable (e.g., conversion of raw test scores to percentile ranks).

Dropout. Dropouts were defined in ELS:2002 as sample members who had been absent from school for 4 or more consecutive weeks at the time of the survey and who were not absent due to accident or illness.

Early Graduate. Early graduates were defined as sample members who had graduated from high school or obtained certification of high school equivalency (e.g., obtained a GED credential) on or before March 15, 2004.

4. SURVEY DESIGN

Target Population
The ELS:2002 base year comprises two primary target populations—schools with 10th grades and 10th-grade students—in the spring term of the 2001–02 school year. There are two slightly different target populations for the first follow-up. One population consists of those students who were enrolled in the 10th grade in 2002. The other population consists of those students who were enrolled in the 12th grade in 2004. The former population includes students who dropped out of school between 10th and 12th grades, and such students are a major analytical subgroup. The target populations of the ELS:2002 second follow-up (2006) were the 2002 sophomore cohort and the 2004 senior cohort. The sophomore cohort consists of those students who were enrolled in the 10th grade in the spring of 2002 and the 12th-grade cohort comprises those students who were enrolled in the 12th grade in the spring of 2004. The sophomore cohort includes students who were in the 10th grade in 2002 but not in the 12th grade in 2004 (i.e., sophomore cohort members but not senior cohort members). The senior cohort includes students who were 12th-graders in 2004 but were not in the 10th grade in U.S. schools in 2002; they were included through a sample freshening process as part of the first follow-up activities.

Sample Design
The sample design for ELS:2002 is similar in many respects to the designs used in the three prior studies of the National Center for Education Statistics (NCES)

Longitudinal Studies Program: the National Longitudinal Study of the High School Class of 1972 (NLS:72), HS&B, and NELS:88. ELS:2002 is different from NELS:88 in that the ELS:2002 base-year sample students are 10th-graders rather than 8th-graders. As in NELS:88, there were oversamples of Hispanics and Asians in ELS:2002. However, for ELS:2002, counts of Hispanics and Asians were obtained from the Common Core of Data (CCD) and the Private School Universe Survey (PSS) to set the initial oversampling rates.

ELS:2002 used a two-stage sample selection process. First, schools were selected with probability proportional to size, and school contacting resulted in 1,220 eligible public, Catholic, and other private schools from a population of approximately 27,000 schools containing 10th-grade students. Of the eligible schools, 752 participated in the study. These schools were then asked to provide 10th-grade enrollment lists. In the second stage of sample selection, approximately 26 students per school were selected from these lists.

Base-Year Survey. The ELS:2002 base-year sample design comprises two primary target populations—schools with 10th grades and sophomores in these schools—in the spring term of the 2001–02 school year. The base-year survey used a two-stage sample selection process. First, schools were selected. These schools were then asked to provide sophomore enrollment lists.

The target population of schools for the ELS:2002 base year consisted of regular public schools, including state Department of Education schools and charter schools, and Catholic and other private schools that contained 10th grades and were in the United States (the 50 states and the District of Columbia). The sampling frame of schools was constructed with the intent to match the target population. However, selected schools were determined to be ineligible if they did not meet the definition of the target population. Responding schools were those schools that had a survey day (i.e., a day when data collection occurred for students in the school). Of the 1,270 sampled schools, there were 1,220 eligible schools and 752 responding schools (67.8 percent weighted response rate). School-level data reflect a school administrator questionnaire, a library media center questionnaire, a facilities checklist, and the aggregation of student data to the school level. School-level data, however, can also be reported at the student level and serve as contextual data for students.

The target population of students for the full-scale ELS:2002 consisted of spring-term sophomores in 2002 (excluding foreign exchange students) enrolled in schools in the school target population. The sampling frames of students within schools were constructed with the intent to match the target population. However, selected students were determined to be ineligible if they did not meet the definition of the target population. Of the 19,220 sampled students, there were 17,590 eligible students and 15,360 participants (87.3 percent weighted response rate). Student-level data consist of student questionnaire and assessment data and reports from students' teachers and parents.

First Follow-up Survey. The basis for the sampling frame for the first follow-up was the sample of schools and students used in the ELS:2002 base-year sample. There are two slightly different target populations for the follow-up. One population consists of those students who were enrolled in the 10th grade in 2002. The other population consists of those students who were enrolled in the 12th grade in 2004. The former population includes students who dropped out of school between 10th and 12th grades, and such students are a major analytical subgroup. Note that in the first follow-up, a student who is defined as a member of the student sample is either an ELS:2002 spring 2002 sophomore or a freshened first follow-up spring 2004 12th-grader.

If a base-year school split into two or more schools, many of the ELS base-year sample members moved en masse to a new school, and they were followed to the destination school. These schools can be thought of as additional base-year schools in a new form. Specifically, a necessary condition of adding a new school in the first follow-up was that it arose from a situation such as the splitting of an original base-year school, thus resulting in a large transfer of base-year sample members (usually to one school, but potentially to more). Four base-year schools split, and five new schools were spawned from these four schools. At these new schools, as well as at the original base-year schools, students were tested and interviewed. Additionally, student freshening was done, and the administrator questionnaire was administered.

Second Follow-up Survey. The target populations of the ELS:2002 second follow-up (2006) were the 2002 sophomore cohort and the 2004 senior cohort. The 2002 sophomore cohort consists of those students who were enrolled in the 10th grade in the spring of 2002, and the 2004 senior cohort comprises those students who were enrolled in the 12th grade in the spring of 2004. The sophomore cohort includes students enrolled in the 10th grade in 2002, but not in the 12th grade in 2004 (i.e., sophomore cohort members, but not senior

cohort members). The senior cohort includes students enrolled in the 12th grade in 2004, but not in the 10th grade in 2002; they were included through a sample freshening process as part of the first follow-up activities.

The second follow-up fielded sample consisted of 16,430 sample members: 14,100 respondents for both the base year and the first follow-up; 1,200 first follow-up nonrespondents who were base-year respondents; 650 base-year nonrespondents who were subsampled in the first follow-up and responded in the first follow-up; 210 base-year or first follow-up questionnaire-incapable members; 170 freshened respondents in the first follow-up; and 100 base-year respondents who were determined to be out of scope in the first follow-up. Once fielded, some members of the sample of 16,430 were determined to be out of scope. There were 460 out-of-scope second follow-up sample members who fell into five basic groups: deceased, out of country, institutionalized/incarcerated, questionnaire incapable/incapacitated, or unavailable for the duration of the 2006 data collection.

High School Transcript Study. Transcripts were collected for all sample members who participated in at least one of the first two student interviews: the base-year interview or the first follow-up interview. These sample members include base-year respondents who were first follow-up nonrespondents and base-year nonrespondents who were first follow-up respondents. Thus, sample members who were dropouts, freshened sample members, transfer students, homeschooled students, and early graduates are included if they were respondents in either of the first two student interviews. Transcripts were also requested for students who could not participate in either of the interviews because of a physical disability, a mental disability, or a language barrier.

Unlike previous NCES transcript studies, which collected transcripts only from the last school attended by sample members, the ELS:2002 transcript study collected transcripts from all base-year schools and the last school attended by sample members who transferred out of their base-year school. Incomplete records were obtained for sample members who had dropped out of school, had fallen behind the modal progression sequence, or were enrolled in a special education program requiring or allowing more than 12 years of schooling. Eighty-six percent of transcript respondents have 4 complete years of high school transcript information.

Data Collection and Processing

The base-year survey collected data from students, parents, teachers, librarians, and school administrators. Self-administered questionnaires and cognitive tests were the principal modes of data collection. Data collection took place primarily during in-school survey sessions conducted by Research Triangle Institute (RTI) field interviewer or team. Base-year data were collected in the spring term of the 2002 school year. A total of 752 high schools participated, resulting in a weighted school response rate of 67.8 percent. A total of 15,360 students participated, primarily in in-school sessions, for an 87.3 percent weighted response rate. Each sampled student's mathematics teacher and English teacher were given a questionnaire to complete. Weighted student-level coverage rates for teacher data were 91.6 percent (indicating receipt of a report from the math teacher, the English teacher, or both). School administrators and library media coordinators also completed a questionnaire (the weighted response rates were 98.5 percent and 95.9 percent, respectively). Questionnaires were mailed to parents, with a telephone follow-up for nonresponders. Student coverage for parent questionnaires was 87.5 percent (weighted). Survey administrators (SAs) completed a facilities checklist at each school. For the first follow-up, overall, about 89 percent (weighted) of the total ELS:2002 sample (comprising both 2002 sophomores 2 years later and 2004 freshened seniors) was successfully surveyed—whether through completion of a student, transfer student, dropout, homeschool, or early graduate questionnaire. For the second follow-up, the sample represents a subset of the combined population of 10th-graders in the spring term of 2002 and 12th-graders in the spring term of 2004. Of the total sample, approximately 15,900 were considered to be eligible for the 2006, among which 14,200 participated, resulting a 88.4 weighted response rate.

Reference dates. In the base-year survey, most questions referred to the students' experience up to the time of the survey's administration in spring 2002. In the follow-ups, most questions referred to experiences that occurred between the previous survey and the current survey. For example, the first follow-up largely covered the period between 2002 (when the base-year survey was conducted) and 2004 (when the first follow-up was conducted).

Data collection. The base-year student data collection began in schools on January 21, 2002, and ended in schools in June 2002; telephone interviews with nonresponding students ended on August 4, 2002. Data collection from school administrators, library media center coordinators, and teachers ended in September

2002. The parent data collection ended on October 17, 2002. The first follow-up in-school data collection occurred between January and June 2004; out-of-school data collection took place between February and August 2004 and included telephone and in-person interviews. The second follow-up data collection was conducted from January to September 2006. To notify sample members about the start of data collection, all sample members and parent(s) were sent a packet which included instructions for the web-based survey.

During the field test of the base-year study, endorsements were secured from organizations felt to be influential in the eyes of the various entities being asked to participate (school administrators, librarians, teachers, students, and parents). Before school recruitment could begin, it was necessary to obtain permission to contact the schools. The Chief State School Officers (CSSOs) of each state (as well as the District of Columbia) were contacted to approve the study for the state. Permission to proceed to the district level was obtained in all 50 states as well as the District of Columbia. Once state approval was obtained, an information package was sent to the District Superintendent of each district/diocese that had sampled schools in the state. Permission to proceed to the school level was received from 693 of the 829 districts/dioceses having eligible sampled schools (83.6 percent). This represented a total of 891 eligible schools with district/diocese permission to be contacted among 1,060 eligible schools affiliated with districts/dioceses (84.1 percent). For public and Catholic schools, school-level contact was begun as soon as district/diocese approval was obtained. For private non-Catholic schools, it was not necessary to wait for higher approval, though endorsements from various private school organizations were sought. The principal of each cooperating school designated a school coordinator to serve as a point of contact at the school and to be responsible for handling the logistical arrangements. The coordinator was asked to provide an enrollment list of 10th-grade students. For each student, the coordinator was asked to give information about sex, race, and ethnicity, and whether the student had an Individualized Education Program (IEP). Dates for a survey day and two make-up days were scheduled. At the same time, staff members were designated to receive the school administrator and library media center questionnaires. Parental consents were obtained. On the survey day at each school, the survey administrator (SA) checked in with the school coordinator and collected any parental permission forms that had come in.

For the base-year and first follow-up surveys, the SA and survey administrator assistant (SAA) administered the student questionnaire and tests via a group administration. The SA and SAA graded the routing tests (see details in the section of “Cognitive test data”) and edited the student questionnaires for completeness. Makeup sessions were scheduled for students who were unable to attend the first session. Interviews were conducted by CATI for students who were unable to participate in the group-administered sessions. The school administrator, teacher, library media center, and parent questionnaires were self-administered; individuals who did not return their questionnaires by mail within a reasonable amount of time were followed up by telephone. The facilities checklist was completed by the SA based on his/her observations in the building on the school’s survey day.

The first follow-up data collection required intensive tracing efforts to locate base-year sample members who, by 2004, were no longer in their 10th-grade schools, but had dispersed to many high schools. In the spring and again in the autumn of 2003, each base-year school was provided a list of ELS:2002 base-year sample members from their school. The school was asked to indicate whether each sample member was still enrolled at the school. For any sample member who was no longer enrolled, the school was asked to indicate the reason and date the student left. If the student had transferred to another school, the base-year school was asked to indicate the name and location of the transfer school. In the fall of 2003, each base-year school was also asked to provide a list of the 12th-graders enrolled at that school, so this information could be used in the freshening process. For students who had left their base-year school, the school was asked to provide contact information to allow for out-of-school data collection during the first follow-up survey period. Telephone data collection began in February 2004. Sample members identified for initial contact by the telephone unit included those no longer enrolled at the base-year school and those who attended base-year schools that did not grant permission to conduct an in-school survey session. Other cases were identified for telephone follow-up after the survey day and all makeup days had taken place at the school that the sample members attended. Some nonresponding sample members were assigned to SAs for field follow-up. A total of 797 sample members were interviewed in the field. An additional 80 field cases were completed either by mailed questionnaire or telephone interview and were withdrawn from the field assignment.

Data collection for the second follow-up was significantly redesigned to include survey modes and procedures that were completely independent of the in-school orientation of the first follow-up survey. An

important aspect of the second follow-up data collection was that high schools were no longer involved in providing assistance with locating sample members. Tracing and sampling maintenance techniques included the following: batch tracing services for updated address information and telephone numbers; updated locating information obtained from student federal financial aid applications; direct contact with sample members and their parents via mail, telephone, or the Internet; intensive tracing efforts by centralized tracing specialists; intensive tracing efforts by field locating specialists in local areas; and tracing students through postsecondary schools applied to or attended, as specified in the 2004 interview. Also, incentive payments were offered to respondents to maximize their participation.

There were three survey modes in the second follow-up: a web-enabled self-administered questionnaire, CATI, and CAPI. Data collection for the second follow-up began on January 25, 2006. For the first 4 weeks, only web and call-in data collection was made available to sample members. After the initial 4 weeks, outbound CATI data collection efforts were undertaken. The primary purpose of the CATI data collection was to complete telephone interviews with sample members when contacted or to set up an appointment to complete the interview. The CATI instrument was virtually identical to the web self-interview. (The only difference was that the CATI version provided an interviewer instruction on each screen to facilitate administration of each item.) CATI interviewers adhered to standardized interviewing techniques and other best practices in administering the interview. To reach sample members who had not yet participated by web or CATI modes, CAPI data collection commenced on April 17 (8 weeks after the start of outbound CATI calling). The approach for CAPI data collection followed the strategy used successfully in B&B:93/2003 and other recent NCES studies. This approach first identified geographic clusters according to the last known zip codes of sample members who could potentially be assigned to CAPI interviewing. Then, based on the distribution of cases by cluster, those that had the highest concentration of cases were staffed with one or more field interviewers. CAPI interviews were conducted on laptop computers via a web-based interface that used personal web server software. To maintain consistency across interviewing modes, the CAPI interview was identical to the CATI interview. CAPI interviewers were allowed to administer the interview over the telephone, which produced conditions even more similar to CATI interviewing.

Data processing. Data processing activities were quite similar for the base-year survey and the first follow-up. An initial check of student documents for missing data was performed on-site by the SA and SAA staff so that data could be retrieved from the students before they left the classroom. If a student neglected to answer a questionnaire item deemed to be critical, the SA/SAA asked the student to complete it after the end of the second-stage test (see details in the section of “Cognitive test data”).

All TELEform questionnaire scans were stored in an Structured Query Language (SQL) server database. CATI data were exported nightly to ASCII files. Cleaning programs were designed to concatenate CATI and TELEform SQL server data into SAS datasets, adjusting and cleaning variables when formats were not consistent. Special attention was focused on this concatenation to verify that results stayed consistent and to rule out possible format problems. Once questionnaire data were concatenated and cleaned across modes and versions, the following cleaning and editing steps were implemented:

- anomalous data cleaning based on a review of the data with the original questionnaire image;
- rule-based cleaning (changes that were made based on patterns in the data rather than on a review of the images);
- hard-coded edits based on changes recommended by a reviewer, if a respondent misunderstood the questionnaire (e.g., respondent was instructed to enter a percentage, but there was strong evidence that the respondent entered a count rather than a percentage); and
- edits based on logical patterns in the questionnaire (e.g., skip pattern relationships between gate and dependent questions).

All respondent records in the final dataset were verified with the Survey Control System (SCS) to spot inconsistencies. Furthermore, the data files served as a check against the SCS to ensure that all respondent information was included in production reports.

Data processing activities for the second follow-up differed from those in the base-year survey and the first follow-up, because respondents could complete a self-administered web questionnaire as an alternative to the survey modes used in previous years. A database was developed in which case/item-specific issues were

reviewed and new values were recorded for subsequent data cleaning and editing.

Item documentation procedures were developed in all waves of data collection to capture variable and value labels for each item. The wording of the question for each item was also provided as part of the documentation. This information was loaded into a documentation database that could export final data file layouts and format statements used to produce formatted frequencies for review. The documentation database also had tools to produce final electronic codebook input files.

Editing. An application was developed in which case/item-specific issues were reviewed and new values were recorded for subsequent data cleaning and editing. Records were selected for review based on one of the following criteria: random selection, suspicious values found during frequency reviews, values out of expected ranges, interviewer remarks, and values not adhering to a particular skip pattern. The review application provided the case/item-level information, the reason for the review, and a link to the scanned image of the questionnaire. Reviewers determined scanning corrections, recommended changes (if respondents had misinterpreted the question), and reviewed items randomly to spot potential problems that would require more widespread review.

The application was built on an SQL server database that contained all records for review and stored the recommended data changes. Editing programs built in SAS read the SQL server database to obtain the edits and applied the edits to the questionnaire data. Questionnaire data were stored at multiple stages across cleaning and editing programs, so comparison across each stage of data cleaning could be easily confirmed with the recommended edits. Raw data were never directly updated, so changes were always stored cumulatively and applied each time a cleaned dataset was produced. This process provided the ability to document all changes and easily fix errors or reverse decisions upon further review.

Editing programs also contained procedures that output inconsistent items across logical patterns within the questionnaire. For example, instructions to skip items could be based on previously answered questions; however, the respondent may not have followed the proper pattern based on the previous answers. These items were reviewed, and rules were written either to correct previously answered (or unanswered) questions to match the dependent items or blank out subsequent items to stay consistent with previously answered items.

Estimation Methods

The general purpose of the weighting scheme was to compensate for unequal probabilities of selection of students into the base-year sample and freshened students into the first follow-up sample and to adjust for the fact that not all students selected into the sample actually participated.

Weighting.

Student level. Two sets of student weights were computed. There is one set of weights for student questionnaire completion; this is the sole student weight that appears in the public-use file and generalizes to the population of spring 2002 sophomores who were capable of completing an ELS:2002 student questionnaire. A second set of weights, for the expanded sample of questionnaire-eligible and questionnaire-ineligible students, appears only in the restricted-use file. This weight sums to the total of all 10^{th}-grade students.

First, the student-level design weight was calculated. The sample students were systematically selected from the enrollment lists at school-specific rates that were inversely proportional to the school's probability of selection. Specifically, the sampling rate for the student stratum within a school was calculated as the overall sampling rate divided by the school's probability of selection. To maintain control of the sample size and to accommodate in-school data collection, the sampling rates were adjusted, when necessary, so that no more than 35 students were selected. A minimum sample size constraint of 10 students was also imposed, if a school had more than 10 tenth-graders. Adjustments to the sampling rates were also made, as sampling progressed, to increase the sample size in certain student strata that were falling short of the sample size targets. The student sampling weight then was calculated as the reciprocal of the school-specific student sampling rate. The student nonresponse adjustment was performed using Generalized Exponential Models (GEMs) to compute the two student nonresponse adjustment factors. For data known for most, but not all, students, the data collected from responding students and weighted hot-deck imputation were used so that there would be data for all eligible sample students.

School level. School weights were computed in several steps. First, a school-level design weight equal to the reciprocal of the school's probability of selection was calculated; second, the school's design weight was adjusted to account for field-test sampling; third, the school weight was adjusted to account for the probability of the school being released. Next, GEMs, which are a unified approach to nonresponse

adjustment, poststratification, and extreme weight reduction, were used. For data known for most, but not all, schools that would be useful to include in the nonresponse adjustment, weighted hot-deck imputation was used so that there would be data for all eligible sample schools.

Scaling. Item Response Theory (IRT) was used to calibrate item parameters for all cognitive items administered to all students. This makes it possible to obtain scores on the same scale for students who took harder or easier forms of the test. IRT also permits vertical scaling of the two grade levels (10^{th} grade in 2002 and 12^{th} grade in 2004). A scale score estimating achievement level was assigned based on the pattern of right, wrong, and omitted responses on all items administered to an individual student. IRT postulates that the probability of correct responses to a set of test questions is a function of true proficiency and of one or more parameters specific to each test question. Rather than merely counting right and wrong responses, the IRT procedure also considers characteristics of each of the test items, such as their difficulty and the likelihood that they could be guessed correctly by low-ability individuals. IRT scores are less likely than simple number-right or formula scores to be distorted by correct guesses on difficult items if a student's response vector also contains incorrect answers to easier questions.

Imputation. In the base-year study, after the editing process (which included logical imputations), the remaining missing values for 14 analysis variables and two ability estimates (reading and mathematics) were statistically imputed. In the first follow-up study, two new variables were selected for imputation: the spring 2004 student ability estimate for mathematics and the spring 2004 student enrollment status. These variables were chosen because they are key variables used in standard reporting and cross-sectional estimation. Most of the variables were imputed using a weighted hot-deck procedure. Additionally, multiple imputations were used for a few variables, including test scores.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

The variance estimation procedure had to take into account the complex sample design, including stratification and clustering. One common procedure for estimating variances of survey statistics is the Taylor series linearization procedure. This procedure takes the first-order Taylor series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. For stratified multistage surveys, the Taylor series procedure requires analysis strata and analysis primary sampling units (PSUs). Therefore, analysis strata and analysis PSUs were created. The impact of the departure of the ELS:2002 complex sample design from a simple random sample design on the precision of sample estimates can be measured by the design effect.

Design effects. The ELS:2002 sample departs from the assumption of simple random sampling in three major respects: student samples were stratified by student characteristics, students were selected with unequal probabilities of selection, and the sample of students was clustered by school. A simple random sample is, by contrast, unclustered and not stratified. Additionally, in a simple random sample, all members of the population have the same probability of selection. Generally, clustering and unequal probabilities of selection increase the variance of sample estimates relative to a simple random sample, and stratification decreases the variance of estimates.

In the ELS:2002 base-year study, standard errors and design effects were computed at the first stage (school level) and at the second stage (student level). The school administrator questionnaire was the basis for the school-level calculations; however, two items from the library questionnaire were also included. For student-level calculations, items from both the student and parent questionnaires were used. Therefore, three sets of standard errors and design effects were computed (school, student, and parent), which is similar to what was done for NELS:88. Each of the three sets includes standard errors and design effects for 30 means and proportions overall and for subgroups.

The student-level base-year design effects indicate that the ELS:2002 base-year sample was more efficient than the NELS:88 sample and the HS&B sample. For means and proportions based on student questionnaire data for all students, the average design effect in ELS:2002 was 2.35; the comparable figures were 3.86 for NELS:88 sophomores and 2.88 for the HS&B sophomore cohort. For all subgroups, the ELS:2002 design effects are smaller, on average, than those for the HS&B sophomore cohort. The smaller design effects in ELS:2002 compared to those for NELS:88 sophomores are probably due to disproportional strata representation introduced by subsampling in the NELS:88 first follow-up. The smaller design effects in ELS:2002 compared to those for the HS&B sophomore cohort may reflect the somewhat smaller cluster size used in the later survey. The ELS:2002 parent-level design effects are similar to the student-level design

effects. For estimates applying to all students, the average design effect was 2.24 for the parent data and 2.35 for the student data. For almost all subgroups, the average design effect was lower for the parent data than for the student data. The school-level design effects reflect only the impact of stratification and unequal probabilities of selection because the sample of schools was not clustered. Therefore, it could be expected that the design effects for estimates based on school data would be small compared to those for estimates based on student and parent data. However, this is not the case, as the school average design effect is 2.76. The reason for this is that the sample was designed to estimate students with low design effects. In addition to stratifying schools, a composite measure of size was used for school sample selection based on the number of students enrolled by race. This is different from the methodology used for NELS:88. The NELS:88 average school design effect in the base year study was considerably lower: 1.82.

The first follow-up design effects are lower for all respondents and for most of the subgroups than the base-year design effects. For the full sample, the design effect for males is the same as in the base year, the design effects for American Indian or Alaska Native and for multiracial respondents are greater than in the base year, and the design effects for the other 14 subgroups are lower than in the base year. For the panel sample, the design effects for American Indian or Alaska Native and for multiracial respondents are greater than in the base year, and the design effects for the other 15 subgroups are lower than in the base year.

The second follow-up design effects are lower for all respondents and for all of the common subgroups used in design effects calculations than the base-year and first follow-up design effects.

Nonsampling Error

Coverage error. In ELS:2002 base-year contextual samples, the coverage rate is the proportion of the responding student sample with a report from a given contextual source (e.g., the parent survey, the teacher survey, or the school administrator survey). For the teacher survey, the student coverage rate can be calculated as either the percentage of participating students with two teacher reports or the percentage with at least one teacher report. The teacher and parent surveys in ELS:2002 are purely contextual. The school-level surveys (school administrator, library media center, facilities checklist) can be used contextually (with the student as the unit of analysis) or in standalone fashion (with the school as the unit of analysis). Finally, test completions (reading assessments, mathematics assessments) are also calculated on a base of the student questionnaire completers, rather than on the entire sample, and thus express a coverage rate. “Coverage” can also refer to the issue of missed target population units in the sampling frame (undercoverage) or duplicated or erroneously enumerated units (overcoverage).

Completed school administrator questionnaires provide 99.0 percent (weighted) coverage of all responding students. Completed library media center questionnaires provide 96.4 percent (weighted) coverage of all responding students. Of the 15,360 responding students, parent data (either by mailed questionnaire or by telephone interview) were received from 13,490 of their parents. This represents a weighted coverage rate of 87.4 percent.

Nonresponse error. Both unit nonresponse (nonparticipation in the survey by a sample member) and item nonresponse (missing value for a given questionnaire/test item) have been evaluated in ELS:2002.

Unit nonresponse. ELS:2002 has two levels of unit response (see table 6) : school response, defined as the school participating in the study by having a survey day on which the students took the test and completed the questionnaires; and student response, defined as a student completing at least a specified portion of the student questionnaire. The final overall school weighted response rate was 67.8 percent, and the final pool 1[1] weighted response rate was 71.1 percent. The final student weighted response rate was 87.3 percent. Because the school response rate was less than 70 percent in some domains and overall, analyses were conducted to determine if school estimates were significantly biased due to nonresponse.

Nonresponding schools (or their districts) were asked to complete a school characteristics questionnaire. The nonresponding school questionnaire contained a subset of questions from the school administrator questionnaire that was completed by the principals of

[1] The sample was randomly divided by stratum into two release pools and a reserve pool. The two release pools were the basic sample, with the schools in the second pool being released randomly within stratum in waves as needed to achieve the sample size goal. Also, the reserve pool was released selectively in waves by simple random sampling within stratum for strata with low yield and/or response rates, when necessary. Each time schools were released from the second release pool or the reserve sample pool, sampling rates were adjusted to account for the non-responding schools and the new schools.

participating schools. (Of the 469 nonresponding eligible sample schools, a total of 437, or 93.2 percent, completed the special questionnaire.)

The school and student nonresponse bias analyses, in conjunction with the weighting adjustments, were not successful in eliminating all bias. However, they reduced bias and eliminated significant bias for the variables known for most respondents and nonrespondents, which were considered to be some of the more important classification and analysis variables. The relative bias decreased considerably after weight adjustments, especially when it was large before nonresponse adjustment, and the relative bias usually remained small after weight adjustments when it was small before nonresponse adjustment.

Student-level nonresponse. For students, although the overall weighted response rate was approximately 87 percent, the response rate was below 85 percent for certain domains, so a student-level nonresponse bias analysis conditional on the school responding was also conducted. Some information on the characteristics of nonresponding students was available from student enrollment lists. On these lists, data were obtained on IEP status, race/ethnicity, and sex. These data were not provided by all schools (in particular, information on IEP status was often missing, and IEP information was typically relevant only for public schools). Consequently, only the school-supplied race/ethnicity and sex data, as well as the school-level data used in the school nonresponse bias analysis, were utilized in conducting the student-level nonresponse bias analysis.

For the student-level nonresponse bias analysis, the estimated bias decreased for every variable after weight adjustments were made. Therefore, the number of significantly biased variables decreased from 42 before adjustment to zero after adjustment.

Item nonresponse. There were no parent or teacher questionnaire items with a response rate that fell below 85 percent. However, there were 78 such items in the student questionnaire, including composites. Item nonresponse was an issue for the student questionnaire because, in timed sessions, not all students reached the final items. The highest nonresponse was seen in the final item, which was answered by only 64.6 percent of respondents.

Table 6. Unit-level and overall weighted response rates for selected ELS:2002 student populations, by data collection wave

Population	Unit-level weighted response rate: Base year school level	Base year student level	1st follow-up	2nd follow-up
Interviewed students	67.8	87.3	93.4	88.4
Tested students	67.8	95.1	87.4	†
Transfers	67.8	†	68.4	81.6
Dropouts	67.8	†	73.2	83.1
	Overall weighted response rate			
		Base year student level	1st follow-up	2nd follow-up
Interviewed students		59.2	63.3	59.9
Tested students		64.5	59.3	†
Transfers		†	46.4	55.3
Dropouts		†	49.6	56.3

† Not applicable.

SOURCE: Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E. (2004). *ELS:2002 Base-Year Data File User's Manual* (NCES 2004-405). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E. (2005). *Education Longitudinal Study of 2002/2004: Base-Year to First Follow-up Data File Documentation* (NCES 2006-344). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Ingels, S.J., Pratt, D.J., Wilson, D., Burns, L.J., Currivan, D., Rogers, J.E., and Hubbard-Bednasz, S. (2007). *Education Longitudinal Study of 2002: Base-Year to Second Follow-up Data File Documentation* (NCES 2008-347). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

At the school level, 41 administrator items had a response rate that fell below 85 percent (ranging from a high of 84.7 percent to a low of 74.6 percent). No library media center questionnaire items fell below the 85 percent threshold, nor did any facility checklist items. While the school-level items can often be used as contextual data with the student as the basic unit of

analysis, these items are also, with the school weight, generalizable at the school level. Therefore, for the school administrator questionnaire, nonresponse rates and nonresponse bias estimates have been produced at the school level. While item nonresponse in the student questionnaire reflects item position in the questionnaire and the inability of some students to reach the final items in a timed session, nonresponse in the school questionnaire must be explained by two other factors: first, the nature of particular items; second, the fact that some administrators completed an abbreviated version of the questionnaire (the high nonresponse items did not appear in the abbreviated instrument).

Measurement error. In the field test, NCES evaluated measurement error in (1) student questionnaire data compared to parent questionnaire data; and (2) student cognitive test data. See *Education Longitudinal Study: 2002 Field Test Report* (Burns et al. 2003).

Parent-student convergence. Some questions were asked of both parents and students. This served two purposes: first, to assess the reliability of the information collected; second, to determine who was the better source for a given data element. These parallel items included number of siblings, use of a language other than English, and parent/child interactions. Additional items on parents' occupation and education, asked in both the parent and student interviews, were also evaluated for their reliability.

Parent-student convergence was low to medium, depending on the item. For example, the convergence on number of siblings is low. Although both parents and students were asked how many siblings the 10th-grader had, the questions were asked quite differently. It is not clear whether the high rate of disagreement is due to parents incorrectly including the 10th-grader in their count of siblings, the inaccurate reporting of "blended" families, or the differences in how the questions were asked in the two interviews. The parent-student convergence on parents' occupation and education was about 50 percent, very similar to those of the NELS:88 base-year interview.

Reliability of parent interview responses. In the field test, the temporal stability of a subset of items from the parent interview was evaluated through a reinterview administered to a randomly selected subsample of 147 respondents. The reinterview was designed to target items that were newly designed for the ELS:2002 interview or revised since their use in a prior NELS interview. Percent agreement and appropriate correlational analyses were used to estimate the response stability between the two interview administrations. The overall reliability of parent interview responses varied from very high to very low, depending on the item. For example, the overall reliability for items pertaining to family composition and race and ethnicity is high; the overall reliability for items pertaining to religious background, parents' education, and educational expectations for the 10th-grader is only marginally acceptable.

Cognitive test data. The test questions were selected from previous assessments: NELS:88, NAEP, and PISA. Items were field tested 1 year prior to the 10th- and 12th-grade surveys, and some items were modified based on field-test results. Final forms were assembled based on psychometric characteristics and coverage of framework categories. The ELS:2002 assessments were designed to maximize the accuracy of measurement that could be achieved in a limited amount of testing time, while minimizing floor and ceiling effects, by matching sets of test questions to initial estimates of students' achievement. In the base year, this was accomplished by means of a two-stage test. In 10th grade, all students received a short multiple-choice routing test, scored immediately by survey administrators who then assigned each student to a low-, middle-, or high-difficulty second-stage form, depending on the student's number of correct answers in the routing test. In the 12th-grade administration, students were assigned to an appropriate test form based on their performance in 10th grade. Cut points for the 12th-grade low, middle, and high forms were calculated by pooling information from the field tests for 10th and 12th grades in 2001, the 12th-grade field test in 2003, and the 10th-grade national sample. Item and ability parameters were estimated on a common scale. Growth trajectories for longitudinal participants in the 2001 and 2003 field tests were calculated, and the resulting regression parameters were applied to the 10th-grade national sample.

The scores are based on IRT, which uses patterns of correct, incorrect, and omitted answers to obtain ability estimates that are comparable across different test forms. In estimating a student's ability, IRT also accounts for each test question's difficulty, discriminating ability, and a guessing factor.

Data Comparability

As part of an important historical series of studies that repeats a core of key items each decade, ELS:2002 offers the opportunity for the analysis of trends in areas of fundamental importance, such as patterns of coursetaking, rates of participation in extracurricular activities, academic performance, and changes in goals and aspirations.

Comparability with NLS:72, HS&B, and NELS:88. The ELS:2002 base-year and first follow-up surveys contained many data elements that were comparable to items from prior studies. *Some items are only approximate matches, and for these, analysts should judge whether they are sufficiently comparable for the analysis at hand. In other cases, question stems and response options correspond exactly across questionnaires.* These repeated items supply a basis for comparison with earlier sophomore cohorts (such as 1980 sophomores in HS&B and 1990 sophomores in NELS:88). With a freshened senior sample, the ELS:2002 first follow-up supports comparisons to 1972 (NLS:72), 1980 (HS&B), and 1992 (NELS:88). The first follow-up academic transcript component offers a further opportunity for cross-cohort comparisons with the high school transcript studies of HS&B, NELS:88, and NAEP.

Although the four studies have been designed to produce comparable results, they also have differences that may affect the comparability as well as the precision of estimates. Analysts should be aware of and take into account these several factors. In particular, there are differences in sample eligibility and sampling rates, in response rates, and in key classification variables, such as race and Hispanic ethnicity. Other differences (and possible threats to comparability) are imputation of missing data, differences in test content and reliability, differences in questionnaire content, potential mode effects in data collection, and possible questionnaire context and order effects.

Eligibility. Very similar definitions were used across the studies in deciding issues of school eligibility. Differences in student sampling eligibility, however, are more problematic. Although the target population is highly similar across the studies (all students who can validly be assessed or, at a minimum, meaningfully respond to the questionnaire), exclusion rules and their implementation have varied somewhat, and exclusion rates are known to differ, where they are known at all. For instance, a larger proportion of the student population was included in ELS:2002 (99 percent) than in NELS:88 (95 percent), which may affect cross-cohort estimates of change.

Sample design. Differences in sampling rates, sample sizes, and design effects across the studies also affect precision of estimation and comparability. Asian students, for example, were oversampled in NELS:88 and ELS:2002, but not in NLS:72 or HS&B, where their numbers were quite small. The base-year (1980) participating sample in HS&B numbered 30,030 sophomores. In contrast, 15,360 sophomores participated in the base year of ELS:2002. Cluster sizes within school were much larger for HS&B (on average, 30 sophomores per school) than for ELS:2002 (just over 20 sophomores per school); larger cluster sizes are better for school effects research, but carry a penalty in greater sample inefficiency. Mean design effect (a measure of sample efficiency) is also quite variable across the studies: for example, for the 10^{th} grade, it was 2.9 for HS&B and 3.9 for NELS:88 (reflecting high subsampling after the 8^{th}-grade base year), with the most favorable design effect, 2.4, for the ELS:2002 base year. Other possible sources of difference between the cohorts that may impair change measurement are different levels of sample attrition over time and changes in the population of nonrespondents.

Imputation of missing data. One difference between the SES variable in ELS:2002 and in prior studies arises from the use of imputation in ELS:2002. Because all the constituents of SES are subject to imputation, it has been possible to create an SES composite with no missing data for ELS:2002. For the HS&B sophomores, SES was missing for around 9 percent of the participants, and for NELS:88 (in 1990) for just under 10 percent.

Score equating. ELS:2002 scores are reported on scales that permit comparisons with reading and mathematics data for NELS:88 10^{th}-graders. Equating the ELS:2002 scale scores to the NELS:88 scale scores was completed through common-item, or *anchor*, equating. The ELS:2002 and NELS:88 tests shared 30 reading and 49 math items. These common items provided the link that made it possible to obtain ELS:2002 student ability estimates on the NELS:88 ability scale. Parameters for the common items were fixed at their NELS:88 values, resulting in parameter estimates for the noncommon items that were consistent with the NELS scale.

Transcript studies. ELS:2002, NELS:88, HS&B, and NAEP were designed to support cross-cohort comparisons. ELS:2002, NAEP, and NELS:88, however, provide summary data in Carnegie units, whereas HS&B provides course totals. In addition, unlike previous NCES transcript studies, which collected transcripts from the last school attended by the sample member, the ELS:2002 transcript study collected transcripts from all base-year schools and the last school attended by sample members who transferred out of their base-year school.

Other factors should be considered in assessing data compatibility. There are some mode-of-administration differences across the studies (for example, ELS:2002 collected 2006 data via self-administration on the Web, as well as by CATI and CAPI; in contrast, NLS:72 and

HS&B used paper-and pencil mail surveys). Order and context effects are also possible (questions have been added, dropped, and reordered, over time).

Comparability with PISA. A feature of ELS:2002 that expands its power beyond that of its predecessors is that it can be used to support international comparisons. Items from PISA were included in the ELS:2002 achievement tests. PISA, which is administered by the Organization for Economic Cooperation and Development, is an internationally standardized assessment, jointly developed by the 32 participating countries (including the United States) and administered to 15-year-olds in groups in their schools. ELS:2002 and PISA test instruments, scoring methods, and populations, however, differ in several respects that impact the equating procedures and interpretation of linked scores.

6. CONTACT INFORMATION

For content information on ELS:2002, contact:

John G. Wirt
Phone: (202) 502-7478
E-mail: john.wirt@ed.gov

Jeffrey A. Owings
Phone: (202) 502-7423
E-mail: jeffrey.owings@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Bozick, R., Lytle, T., Sigel, P.H., Ingels, S.J., Rogers, J.E., Lauff, E., and Planty, M. (2006). *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation (Restricted Use)* (NCES 2006-338). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Cahalan, M.W., Ingels, S.J., Burns, L.J., Planty, M., and Daniel, B. (2006). *United States High School Sophomores: A Twenty-Two Year Comparison, 1980-2002* (NCES 2006-327). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S.J., Burns, L.J., Charleston, S., Chen, X., and Cataldi, E.F. (2005). *A Profile of the American High School Sophomore in 2002: Initial Results From the Base Year of the Longitudinal Study of 2002* (NCES 2005-338). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E. (2004). *ELS: 2002 Base-Year Data File User's Manual* (NCES 2004-405). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E. (2005). *Education Longitudinal Study of 2002/2004: Base-Year to First Follow-up Data File Documentation* (NCES 2006-344). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S.J., Pratt, D.J., Wilson, D., Burns, L.J., Currivan, D., Rogers, J.E., and Hubbard-Bednasz, S. (2007). *Education Longitudinal Study of 2002: Base-Year to Second Follow-up Data File Documentation* (NCES 2008-347). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Uses of Data

Bozick, R., Lytle, T., Siegel, P.H., Ingels, S.J., Rogers, J.E., Lauff, E., and Planty, M. (2006). *Education Longitudinal Study: 2002/2004 Restricted-Use Base-Year, First Follow-up, and High School Transcript Data Files and Electronic Codebook System* (NCES 2006-351). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E. (2004). *Education Longitudinal Study: 2002 Data Files and Electronic Codebook System* (NCES 2004-404). National Center for Education

Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E. (2005). *Education Longitudinal Study: 2002/2004 Restricted-Use Base-Year to First Follow-up Data Files and Electronic Codebook System* (NCES 2006-330). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E. (2005). *Education Longitudinal Study: 2002/2004 Public Use Base-Year to First Follow-up Data Files and Electronic Codebook System* (NCES 2006-346). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Burns, L.J., Heuer, R., Ingels, S.J., Pollack, J., Pratt, D.J., Rock, D., Rogers, J., Scott, L.A., Siegel, P., and Stutts, E. (2003). *Education Longitudinal Study: 2002 Field Test Report* (NCES Working Paper 2003-03). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 10: SASS School Library Media Center Survey (SLS)

1. OVERVIEW

SAMPLE SURVEY OF ELEMENTARY AND SECONDARY SCHOOL LIBRARIES

The School Library Media Center Survey collects data on:

- Collections
- Expenditures
- Technology
- Services

Federal surveys of school library media centers in elementary and secondary schools in the United States were conducted in 1958, 1962, 1974, 1978, and 1985. The National Center for Education Statistics (NCES) asks questions about libraries in public and Bureau of Indian Education (BIE) schools as part of the Schools and Staffing Survey (SASS) (see chapter 4 for details on SASS). The School Library Media Center Survey was introduced as a component of SASS in the 1993–94 school year. The survey was administered to both public and private schools in the 1993–94 and 1999–2000 SASS, but only to public schools (including BIE-funded schools) in the 2003–04 and 2007–08 SASS. It is sponsored by NCES and administered by the U.S. Bureau of the Census.

Purpose

The purpose of the School Library Media Center Survey is to provide a national picture of school library collections, expenditures, technology, and services. The survey furnishes national estimates for public school libraries (by school grade level and urbanicity) and for libraries operated by BIE schools; state estimates are also provided for public school libraries.

Components

The School Library Media Center Survey was introduced in the 1993–94 SASS.

The 1993–94 School Library Media Center Survey consisted of two components, one on the school's library media center and the other on the library media specialist. The 1999–2000, 2003–04, and 2007–08 SASS administrations included only the library media center component. The survey is sent to public schools, including BIE schools, in the 50 states and the District of Columbia. Until the 2003–04 SASS, the survey was also sent to private schools.

School Library Media Center Survey. The library survey is designed to provide a national picture of school library media center facilities, collections, equipment, technology, staffing, income, expenditure, and services. A section on information literacy was added to the 2003–04 and 2007–08 surveys. The respondents are school librarians or other school staff members familiar with the library.

The School Library Media Center survey was designed to profile the school library media specialist workforce, including demographic characteristics, academic background, workload, career histories and plans, compensation, and perceptions of the school library media specialist profession and workplace. The eligible respondent was the staff member whose main assignment at the school was to oversee the library.

Periodicity

The library survey was conducted in the 1993–94, 1999–2000, 2003–04, , 2007–08 SASS, and will be conducted again in 2011-12. The 1993 and 1999-2000

collections covered public, private, and BIE schools; collections since then covered only public and BIE schools.

2. USES OF DATA

School libraries and library media centers are an important component of the educational process. Data from the library survey provide a national picture of school library collections, expenditures, technology, and services. The information can be used by federal, state, and local policymakers and practitioners to assess the status of school library media centers in the United States. It also contributes to the assessment of the federal role in supporting school libraries. The librarian survey provided, for the first time, a national profile of the school library media specialist/librarian workforce.

These data can also be used to address current issues related to school libraries. Recent interest has focused on the contribution that libraries could make to the current education reform movement. Education reform has prompted increased attention to the role that school libraries/media centers might play in applying new technology and developing new teaching methods. Some analysts argue that libraries have a crucial role in developing computer literacy and educating students in the use of modern information technologies. A number of observers also have argued that expanding the function of libraries is a key prerequisite to meeting the National Education Goals.

3. KEY CONCEPTS

Some of the key concepts and terms in the School Library Media Center Survey are defined below.

Librarian. A school staff member whose main responsibility is taking care of the library.

Library Media Center. An organized collection of printed, audiovisual, or computer resources that (a) is administered as a unit, (b) is located in a designated place or places, and (c) makes resources and services available to students, teachers, and administrators.

Library Media Specialist. A teacher who is state certified in the field of library media.

4. SURVEY DESIGN

Target Population

The universe of library media centers/ libraries in elementary and secondary schools with any of grades 1–12 in the 50 states and the District of Columbia.

Sample Design

In 1993–94, the library media center sample was a subsample of the SASS school sample. Drawn from the 13,000 schools in the SASS, the library sample consisted of 5,000 public schools, 2,500 private schools, and the 180 BIE schools in the United States.

The strata used for library sampling were the same as those used in the public school sampling of the Schools and Staffing Survey (SASS) (see chapter 4 for details) (state and grade level). All BIE schools were selected for the library survey, so no stratification or sorting was needed. Within strata, public schools in the 1993–94 sample were sorted on the following variables:

- local education agency (LEA) metro status: 1 = central city of a metropolitan statistical area (MSA), 2 = MSA (not central city), 3 = outside MSA;
- Common Core of Data (CCD) LEA ID;
- school enrollment; and
- CCD school ID.

The sample schools were then systematically subsampled using a probability proportionate to size algorithm, where the measure of size was the square root of the number of teachers in the school as reported in the CCD (the public school sampling frame for SASS) multiplied by the inverse of the school's probability of selection from the public school sample file. Any school with a measure of size larger than the sampling interval was excluded from the library sampling operation and included in the sample with certainty.

The private school library frame for 1993–94 was identical to the frame used for the SASS private school survey, except that it excluded schools with special program emphasis (special education, vocational, or alternative curriculum schools). Private schools were stratified by recoded affiliation (Catholic, other religious, nonsectarian); grade level (elementary, secondary, combined); and urbanicity (urban, suburban, rural). Within each stratum, sorting occurred

on the following variables: (1) frame (list frame and area frame); and (2) school enrollment.

Within each stratum, private schools were systematically selected using a probability proportionate to size algorithm. The measure of size used the school's measure of size (i.e., the square root of the number of teachers in the school as reported in the CCD) multiplied by the inverse of the school's probability of selection. Any library with a measure of size larger than the sampling interval was excluded from the probability sampling process and included in the sample with certainty. In all, 2,500 private schools were selected for the library sample in the 1993–94 SASS. In 1999-2000, the Library Media Center questionnaire was administered to all school within the SASS sample.

In 2003–04 and 2007-08, the Library Media Center questionnaire was administered to public and BIE SASS school samples, excluding private schools. Each sampled school received a library media center questionnaire. The sampling design for Library Media Center Survey follows that of the public school sample and BIE school sample of SASS. The BIE schools were selected for the sample with certainty. A number of changes were made in the sample design (i.e., stratification, sample sizes, sample sort, and school definition) from the 1999-2000 SASS to the 2003-04 SASS to the 2007-08 SASS. See more information on the 2007-08 public and BIE school sampling in Schools and Staffing Survey (SASS) (chapter 4).

Data Collection and Processing

The U.S. Bureau of the Census is the collection agent for the School Library Media Center Survey. Data collection and processing procedures are discussed below.

Reference dates. Most data items refer to the most recent full week in the current school year. Questions on collections and expenditures refer to the previous school year.

Data collection. The School Library Media Center Survey is delivered with other SASS components beginning in October of the survey year. The survey is delivered to the school librarian or another staff member familiar with the library. (The follow-up procedures are described in chapter 4.)

Editing. Once data collection is complete, data records are processed through a clerical edit, preliminary interview status recode (ISR) classification, computer pre-edit, range check, consistency edit, and blanking edit. (See chapter 4 for details.) After the completion of these edits, records are processed through an edit to make a final determination of whether the case is eligible for the survey and, if so, whether sufficient data have been collected for the case to be classified as an interview. A final ISR value is assigned to each case as a result of the edit.

Estimation Methods

Weighting. In the SASS School Library Media Center component, data are used to estimate the characteristics of schools with library media centers as well as schools without library media centers. Whenever possible, sampled schools with library media centers and sampled schools without library media centers are adjusted separately. Thus, interviewed library media centers are weighted up to the weighted estimate of sampled schools known to have library media centers, as determined at the time school library media center questionnaires were distributed. Likewise, the number of interviewed schools without library media centers is weighted up to the weighted number of all schools without library media centers as determined from the questionnaire distribution. This is done to study the characteristics of each type of school.

When it is not possible to adjust the library weights by the type of school, all sampled school library media centers and schools without library media centers are adjusted as a whole. This is necessary to handle instances in which the existence of the library media center cannot be established during data collection. Due to reporting inconsistencies between the school library media center questionnaire and the school questionnaire, school library media center survey data are not adjusted directly to schools reporting to have library media centers on the school questionnaire.

Imputation. Items from the SASS School Library Media Center questionnaire that still had items that were "not answered" went through a first stage of imputation in which unanswered items were imputed from other items on the same library media center record or items on the corresponding school record. The library media center data then went through the second stage of imputation in which some of the remaining "not answered" items were filled using either the data record from a similar record, regression imputation, or random ratio imputation. The third stage of imputation filled in the remaining "not answered" items that were not resolved during the first two stages of imputation (i.e., imputed clerically). After all stages of imputation were completed and no more "not answered" items remained, the library media center data from BIE-funded schools were separated into a single dataset.

Recent Changes

The School Library Media Center Survey has not been administered to private schools since the 1999-2000 school year.

5. DATA QUALITY AND COMPARABILITY

Although data are imputed for nonrespondents, caution should be exercised when analyzing data by state, sector, or affiliation. Since nonresponse varies by state, the reliability of state estimates and comparisons are affected. Users should be especially cautious about using data at a level of detail where the nonresponse rate is 30 percent or greater. See below for more information on the types of errors affecting data quality and comparability.

Sampling Error

The estimators of sampling variances for SASS statistics take the SASS complex sample design into account. (See chapter 4.)

Nonsampling Error

Nonresponse error.

Unit nonresponse. The weighted unit response rates for the 2007–08 School Library Media Center Survey were 76.9 percent for public schools and 82.1 percent for BIE schools.

Item nonresponse. Some 95 percent of the items in the public school version of the 2007–08 School Library Media Center Survey had response rates above 85 percent and 93 percent of the items in the BIE version had response rates above 85 percent. All items in both versions had response rates above 70 percent, and there was no substantial evidence of bias in either case.

Measurement error. A reinterview was conducted for the 2003–04 SASS, but it did not include questions from the School Library Media Center Survey.

6. CONTACT INFORMATION

For content information on the School Library Media Center Survey, contact:

Kerry Gruber
Phone: (202) 502-7349
E-mail: kerry.gruber@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Gruber, K., Rohr, C.L., and Fondelier, S.E. (1996). *1993–94 Schools and Staffing Survey: Data File User's Manual, Volume I: Survey Documentation* (NCES 96-142). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Tourkin, S.C., Pugh, K.W., Fondelier, S.E., Parmer, R.J., Cole, C., Jackson, B., Warner, T., Weant, G., and Walter, E. (2004). *1999–2000 Schools and Staffing Survey (SASS) Data File User's Manual* (NCES 2004-303). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourkin, S.C., Warner, T., Parmer, R., Cole, C., Jackson, B., Zukerberg, A., Cox, S., and Soderborg, A. (2007). *Documentation for the 2003–04 Schools and Staffing Survey* (NCES 2007-337). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tourkin, S., Thomas, T., Swaim, N., Cox, S., Parmer, R., Jackson, B., Cole, C., and Zhang, B. (2010). *Documentation for the 2007–08 Schools and Staffing Survey* (NCES 2010-332). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Uses of Data

Dickson, G. (1998). *Evaluation of Definitions and Analysis of Comparative Data for the School Library Statistics Program* (NCES 98-267). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Survey Design

Abramson, R., Cole, C., Fondelier, S., Jackson, B., Parmer, R., and Kaufman, S. (1996). *1993–94 Schools and Staffing Survey: Sample Design and Estimation* (NCES 96-089). National Center for

Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Feindt, P.J. (1996). *Reinterview Report: Response Variance in the 1993 Library Survey*. U.S. Department of Commerce. Washington, DC: U.S. Bureau of the Census.

Chapter 11: Academic Libraries Survey (ALS)

1. OVERVIEW

The Academic Libraries Survey (ALS) is designed to provide concise information on library resources, services, and expenditures for all academic libraries in the 50 states, the District of Columbia, and the outlying areas. The ALS was conducted by NCES on a 3-year cycle between 1966 and 1988. Between 1988 and 1998, the ALS was a component of the Integrated Postsecondary Education Data System (IPEDS) (see chapter 12 for more details on IPEDS) and was collected on a 2-year cycle. Since 2000, the Academic Libraries Survey has been conducted independently of IPEDS; however, it remains on a 2-year cycle.

ALS collects data biennially from approximately 3,800 degree-granting postsecondary institutions in order to provide an overview of academic libraries nationwide and by state. The 1996 ALS also surveyed libraries in nonaccredited institutions that had a program of 4 years or more. Because so few of these libraries responded to ALS, their data were not published. Beginning with the 1998 ALS, the major distinction has been whether or not the library is part of a postsecondary institution that is eligible for Title IV funds.

Although ALS was a component of IPEDS from 1988 through 1998, beginning in 2000, ALS began collecting data independently of the IPEDS data collection. However, data from ALS can still be linked to IPEDS data using the institution's UNITID number. IPEDS serves as the frame, or universe, of degree-granting postsecondary institutions from which eligible institutions are selected for the current ALS administration.

BIENNIAL SURVEY OF THE UNIVERSE OF LIBRARIES IN POSTSECONDARY EDUCATION INSTITUTIONS

ALS collects data on:

- Library staffing
- Operating expenditures
- Total volumes
- Circulation, loan, and reference transactions
- Electronic services
- Gate count

Purpose
To periodically collect and disseminate descriptive data on all postsecondary academic libraries in the United States, the District of Columbia, and the outlying areas, for use in planning, evaluation, and policymaking.

Components
There is a single component to the Academic Libraries Survey. The survey is completed by a designated respondent at the library. While ALS was a part of IPEDS, an appointed state IPEDS Data coordinator collected the information from academic librarians and submitted it to NCES.

Academic Libraries Survey. An academic library is the library associated with a degree-granting institution of higher education. Academic libraries are identified by the postsecondary institution of which they are a part of (see *Key Concept*s below for further detail). Through 1996, ALS distinguished between libraries in postsecondary institutions accredited by agencies recognized by the Secretary of the U.S. Department of Education and libraries in nonaccredited institutions that had

programs of 4 or more years. Starting with the 1998 collection, the major distinction has been whether or not the library is part of a postsecondary institution that is eligible for Title IV funds.

Data are collected on the number of libraries, branches, and service outlets; full-time-equivalent (FTE) library staff by position; operating expenditures by purpose, including salaries and fringe benefits; total volumes held at the end of the fiscal year; circulation transactions, interlibrary loan transactions, and information services for the fiscal year; hours open, gate count, and reference transactions per typical week; and, as of 1996, the availability of electronic services, such as electronic catalogs of the library's holdings, electronic full-text periodicals, internet access and instruction on use, library reference services by e-mail, electronic document delivery to patrons' account address, computers and software for patron use, scanning equipment for patron use, and services to the institution's distance education students. In 2004, a new set of questions on "information literacy" was added to the questionnaire. In 2010, reference transactions was broken out into "in-person" and "virtual" and "over 20 minutes" and "under 20 minutes." Also, a new set of yes/no questions about "virtual reference" was added to the questionnaire.

Periodicity

Biennial in even-numbered years since 1990; triennial from 1966 through 1988.

2. USES OF DATA

Effective planning for the development and use of library resources demands the availability of valid and reliable statistics on academic libraries. ALS provides a wealth of information on academic libraries. These data are used by federal program staff to address various policy issues, by state policymakers for planning and comparative analysis, and by institutional staff for planning and peer analysis. Specific uses are listed below:

- Congress uses ALS data to assess the impact of library grants programs, the need for revisions to existing legislation, and the allocation of funds.

- Federal agencies that administer library grants for collections development, resource sharing, and networking activities require ALS data for their evaluation of the condition of academic libraries.

- State education agencies use ALS data to make comparisons at the national, regional, and state levels.

- Accreditation review programs for academic institutions require current library statistical data in order to evaluate postsecondary education institutions, establish standards, and modify comparative norms for assessing the quality of programs.

- Library administrators, academic managers, and national postsecondary education policy planners need current data on new electronic technologies to assess the impact of rapid technological change on the collections, budgets, and staffs of academic libraries. College librarians and administrators need these data to develop plans for the most effective use of local, state, and federal funds. Staff data are input to supply/demand models for professional and paraprofessional librarians.

- Library associations—such as the American Library Association, the Association of Research Libraries, and the Association of College and Research Libraries—use ALS data to determine the general status of the profession. Other research organizations use the data for studies of libraries.

- Program staff in the Institute of Education Sciences of the U.S. Department of Education use ALS data for administering their library grants program, evaluating existing programs, and preparing documentation for congressional budget hearings and inquiries.

3. KEY CONCEPTS

Some of the key concepts and terms in ALS are defined below. For additional terms, refer to *Documentation for the Academic Library Survey (ALS) Public Use Data File: 2008* (Phan, Hardesty, and Sheckells, 2009).

Academic Library. An entity in a postsecondary education institution that provides all of the following: (1) an organized collection of printed or other materials, or a combination thereof; (2) a paid, trained library staff to provide and interpret library materials to meet the informational, cultural, recreational, or educational needs of clientele; (3 an established hours

of operation during which paid, trained staff are available to meet the informational service needs of clientele; and (4) the physical facilities necessary to support such a collection, staff, and schedule. This definition includes libraries that are part of learning resource centers.

Branch Library. An auxiliary library service outlet with quarters separate from the central library of an institution. A branch library has a basic collection of books and other materials, a regular staffing level, and an established schedule. Branch libraries are administered either by the central library, as in the case of some libraries (such as law or medical libraries), or through the administrative structure of other units within the university. Departmental study/reading rooms are not included. Libraries on branch campuses that have separate NCES identification numbers are reported as separate libraries.

Child Institution. A “child” institution does not respond directly to the ALS or IPEDS data collections. The data for such an institution are aggregated with and reported by its “parent” institution.

Volume. Any printed, mimeographed, or processed work, contained in one binding or portfolio, hardbound or paperbound, that has been cataloged, classified, or otherwise made ready for use.

Title. A publication that forms a separate bibliographic whole (whether issued in one or several volumes, reels, disks, slides, or parts). The term applies equally to printed materials (e.g., books and periodicals), sound recordings, film and video materials, microforms, and computer files.

Circulation Transaction. Includes all items lent from the general collection and from the reserve collection for use generally (although not always) outside the library. Includes both activities with initial charges (either manual or electronic) and renewals, each of which is reported as a circulation transaction.

Interlibrary Loan. A transaction in which library materials, or copies of the materials, are made available by one library to another upon request. Loans include providing materials and receiving materials. Libraries involved in these interlibrary loans cannot be under the same administration or on the same campus.

Reference Transaction. These are information contacts that involve the knowledge, use, recommendation, interpretation, or instruction in the use of one or more information sources by a member of the library staff. Information sources may include printed (e.g., book volumes) and nonprinted (e.g., microforms) materials and machine-readable databases (e.g., those on CD-ROM). The transaction may include providing direction to services outside the library.

Online Public Access Catalog (OPAC). A library’s catalog of its collections in electronic form, accessible by computer or other online workstation.

Gate Count. The total number of persons physically entering the library in a typical week. A single person can be counted more than once.

4. SURVEY DESIGN

Target Population

The libraries of all institutions in the 50 states, the District of Columbia, and the outlying areas that have as their primary purpose the provision of postsecondary education. Branch campuses of U.S. institutions located in foreign countries are excluded. Through 1996, ALS distinguished between libraries in postsecondary institutions accredited by agencies recognized by the Secretary of the U.S. Department of Education and libraries in nonaccredited institutions that had programs of 4 or more years. In 1996, there were approximately 3,600 accredited institutions and 400 nonaccredited institutions in the IPEDS universe. About 3,400 of the accredited institutions had academic libraries. Starting with the 1998 collection, the major distinction has been whether or not the library is part of a postsecondary institution that is eligible for Title IV funds. In 2004, there were 3,700 Title IV eligible, degree-granting postsecondary institutions in the 50 states and the District of Columbia that had academic libraries. In 2006, the reported number of the nation’s Title IV eligible institutions with academic libraries was 3,600. In 2008, the reported number of the nation’s Title IV eligible institutions with academic libraries was 3,800.

Sample Design

ALS surveys the universe of postsecondary institutions.

Data Collection and Processing

For the 1990, 1992, 1994, 1996, and 1998 data collections, state IPEDS Data coordinators collected, edited, and submitted ALS data to the U.S. Census Bureau, using the software package Input and Data Editing for Academic Library Statistics (IDEALS). An academic librarian in the state assisted with the collection and submission of the data.

Since 2000, ALS has not been a component of the IPEDS survey system. The 2000, 2002, 2004, 2006, and 2008 ALS surveys were web collections. The U.S. Census Bureau is the collection agent. State-level library representatives are available to promote responses from librarians and to assist in problem resolution when anomalies are discovered in responses.

Reference dates. Most ALS data are reported for the most recently completed fiscal year, which generally ends before October 1 of the survey year. Information on staff and services per typical week are collected for a single point in time during the fall of the survey year, usually the institution's official fall reporting date or October 15.

Data collection. In the 2000, 2002, 2004, 2006, and 2008 ALS data collections, library respondents submitted data directly to the Census Bureau through the Web. For the 2008 web-based data collection, state-level library representatives were available to promote prompt responses from librarians. A web-based survey is the latest in a number of steps to improve ALS collection.

In July 1990, NCES initiated an ALS improvement project with the assistance of the National Commission on Libraries and Information Science (NCLIS) and the American Library Association's Office of Research and Statistics (ALA-ORS). The project identified an academic librarian in each state to work with the IPEDS coordinators in submitting their library data. During the 1990s, many of these library representatives took the major responsibility for collecting data in their state. Others were available to assist in problem resolution when anomalies were discovered in completed questionnaires.

The ALS improvement project also led to the development of the microcomputer software package IDEALS, which was used by states in reporting their academic library data from 1990 through 1998. Along with the software, NCES provided state IPEDS Data coordinators with a list of instructions explaining precisely how responses were to be developed for each ALS item. Academic librarians within each state completed hard-copy forms, as they had previously, and returned them to the state's library representative or IPEDS coordinator. States were given the option of submitting the paper forms, but were encouraged to enter the data into IDEALS and submit the data on diskette to the Census Bureau. Nearly all states elected the diskette option.

ALS was mailed to postsecondary institutions during the summer of the survey year, with returns requested during the fall. Any survey returns from institutions that did not have an academic library were declared to be out of scope, as were institutions that did not have their own library but shared one with other institutions. In recent years, less than half of the nonaccredited institutions responded to the survey; NCES does not include data on this group in publications.

Editing. The web-based data collection collection application features internal edit checks. An edit check tool alerts the respondent to questionable data via interactive "edit check warnings" during the data entry process and through edit check reports that can be viewed on screen or printed. The edit check program enables the respondent to submit an edited data to NCES which usually required little or no follow-up for data problems. The edit check tool includes seven types of edits: Summations, Relational edit checks, Range checks, Current year/prior year comparisons, ratios, item comparison, and missing or blank items.

After responses are received, the U.S. Census Bureau reviews the data and contacts respondents with questionable data to request verification or correction of that data. Data records are then aggregated into preliminary draft tables, which are reviewed by NCES and the U.S. Census Bureau for data quality issues. Once all edits have been performed and all corrections have been made, the data undergo imputation to compensate for nonresponse (see below). (For more information on the edit check, please see appendix A in Phan, Hardesty, Sheckells, and Davis [2009])

Estimation Methods

Imputation is used in ALS to compensate for nonresponse. In 1994, the procedures were changed to use data from the previous survey, if available, and to only use imputation group means (see below) if prior-year data were not available. Before 1994, only imputation group medians were used.

Imputation. ALS imputation is based on the response in each part of the survey. Most parts go through either total or partial imputation procedures, except for the following items: (1) Number of branch and independent libraries; (2) Library staff information – contributed services staff; and (3) Library operating expenditures – employee fringe benefits. These items are imputed only if reported prior-year data are available (contributed services staff and employee fringe benefits apply to only a few institutions). Items (1) Electronic Services, and (2) Information Literacy do not go through imputation.

The imputation methods use either prior-year data or current-year imputation group means. The procedures

are slightly different depending on whether an institution is totally nonresponding or partially nonresponding in the current year. If prior-year data are available, the imputation procedure either carries forward the prior-year data or carries forward the prior-year data multiplied by a growth factor. If prior-year data are not available, the imputation procedure uses the current-year imputation group medians (in the 2002, 2004, 2006, and 2008 ALS) or means (in previous survey cycles) as the imputed value.

Medians/means and ratios are calculated for each of the imputation groups (27 imputation groups in the 2008, 2006, 2004, and 2002 ALS and 8 imputation groups in the 2000 and 1998 ALS). In2008, 2006, 2004 and 2002, the imputation cells were determined based on sector and FTE enrollment. The sector categories used were (1) public, 4-year or above; (2) private nonprofit, 4-year or above; (3) private for-profit, 4-year or above; (4) public, 2-year; (5) private nonprofit, 2-year; and (6) private for-profit, 2-year. The use of FTE to determine imputation cells was not employed until 2002. In 1998 and 2000, the strata were based upon the highest level of degree (doctor's, master's, bachelor's, and associate's) and control and size of institution. The four control/size imputation categories were (1) public, less than median number of degrees for institutions in that category; (2) public, equal to or greater than the median; (3) private, less than the median; and (4) private, equal to or greater than the median. Note that computation of the imputation base excludes institutions that merged, split, submitted combined forms, changed sectors from the prior year, or did not submit a full report for either the current or prior year.

After imputation, if a total was missing or known to need adjustment, then the total was readjusted to equal the sum of its detail items.

Using a ratio adjustment to prior-year data represented a change from the imputation procedures followed in cycles prior to 1996, and may have resulted in some small differences in estimates. While checks indicate that the effect of the change was not large, caution should be exercised in making comparisons with pre-1996 or earlier reports (see Cahalan, Mansfield, and Justh 2001). Using FTE to determine imputation cells and using medians instead of means for imputation also represents a change from the procedures followed in cycles prior to 2002. While research indicates that the effect of the change in imputation procedure was not large, caution should be exercised in making comparisons with reports from 2000 or earlier (see Phan, Hardesty, and Sheckells 2009).

Recent Changes

Before 2000, ALS was a component of IPEDS; the state IPEDS Data coordinators collected, edited, and submitted ALS data to the Census Bureau, using the software package IDEALS. Since 2000, ALS data have been collected over the Internet via a web-based reporting system. The Census Bureau is the collection agent.

Several changes were made to the survey instrument in 1996, 1998, 2000, 2002, 2004, 2006, 2008, and 2010. These are summarized below.

In the 1996 instrument, the data items in part E (Library Services) were expanded to request separate reporting for returnable and nonreturnable, as well as totals. In addition, a new section, part G, was added to collect information about access to the following electronic services, both on and off campus:

- electronic catalog that includes the library's holdings;
- electronic indexes and reference tools;
- electronic full-text periodicals;
- electronic full-text course reserves;
- electronic files other than the catalog (e.g., finding aids, indices, manuscripts) created by library staff;
- Internet access;
- library reference service by e-mail;
- capacity to place interlibrary loan/document delivery requests electronically;
- electronic document delivery by the library to patrons' account/address;
- computers not dedicated to library functions for patron use inside the library;
- computer software for patron use inside the library (word processing, spreadsheet, custom applications, etc.);
- technology in the library to assist patrons with disabilities (TDD, specially equipped workstations, etc.); and
- instruction by library staff on the use of internet resources.

The 1998 ALS survey instrument modifications included the following.

The definition of a library was moved to the cover page and reformatted as a checklist. The other cover page change was that the possibilities of reporting data for another library or having data reported by another library were clarified. The data items in part B (Library Staff) were expanded to request a total FTE count for librarians and other professionals as well as separate counts of these two categories of staff. Part C was renamed “Library Expenditures” and the word “operating” was used only in reference to expenditures for items other than staff and materials. The two major lines for reporting expenditures on information resources were subdivided as follows: books, serial backfiles, and other materials (paper and microform; electronic); and current serial subscriptions and search services (paper and microform; electronic). In addition, expenditures on search services were to be reported with those for current serial subscriptions, in recognition of the fact that it is often impossible to separate the two.

Part D (Collections) was changed the most, being reduced from 18 to 7 lines. It collected data on only three types of materials: books, serial backfiles, and other materials (paper; microform; electronic); current serial subscriptions (paper and microform; electronic); and audiovisual materials. The following lines were deleted: manuscripts and archives, cartographic materials, graphic materials, sound recordings, film and video materials, and computer files. Except for paper materials, there was no longer separate reporting of physical counts and title counts. In part F (Library Services, Typical Week), “public service hours” was changed to “hours open” since some libraries keep two separate counts and are unsure of what to report. “Typical week” was added to the heading above the space for reporting figures to reinforce that only typical week figures should be reported.

In part G (Electronic Services), the following items were added to the yes/no checklist about access to electronic services:

- computers not dedicated to library functions for patron use inside the library;
- computer software for patron use in the library (word processing, spreadsheet, custom applications, etc.);
- scanning equipment for patron use in the library; and
- services to your institution’s distance education students.

The changes to the 2000 ALS form were as follows:

Cover sheet (Library Definition): The format of the question regarding providing financial support to another library was clarified.

Part C (Library Expenditures): The text for library expenditures was modified to clarify what is wanted.

Part D (Library Collections): The items “Electronic - Titles” and “Number of electronic subscriptions” were dropped and the item covering other forms of subscriptions was revised.

Part E (Library Services): A new item was added for “documents delivered from commercial services,” and the words “document delivery” were dropped from the items for “interlibrary loans provided” and “interlibrary loans received.” The item on “reserve collections” was dropped and the preceding line was revised to read “Circulation Transactions (including reserves).”

Part G (Electronic Services): Five items were added under the heading “Consortial Services.”

The 2002 ALS survey instrument underwent the following changes:

Part B (Library Staff): A new column 2 was added, “Salaries and Wages – library expenditures for staff” (previously in part C); number of full-time equivalents (FTE) became column 1; contributed services staff was dropped; and fringe benefits were added.

Part C (Library Expenditures):

- Change in wording for the note at the top of page: “Do not report the same expenditures more than once” was removed. New wording: “See instructions for exclusions and definitions.”
- Breakout of staff salaries and wage expenditures moved to part B.
- Total salaries and wages line added to library expenditures.
- Line 10—books, serial backfiles, and other materials (one-time purchases)—became a total line.

- Electronic and audiovisual lines became subsets of the line 10 total.

- Line 13—current serial subscriptions (ongoing commitments)—became a total line.

- Electronic serials line became a subset of the line 13 total.

- "Other materials" changed to "Other expenditures for information resources."

- Furniture and equipment line was dropped and was included in "all other operating expenditures," line 20.

- Fringe benefits lines moved to part B.

Part D (Collections): "Paper – volumes" changed to "Books, serial backfiles, and other paper materials (including government documents)"; dropped line for paper titles; added line for E-books; reversed the sequence of the next two lines: "Current serial subscriptions" and "Audiovisual materials."

Part E (Library Services): Divided circulation into general and reserve by having two lines as follows: line 34a – "General circulation transactions"; and line 34b – "Reserve circulation transactions."
Part G (Electronic Services): Parts G1 and G2 were combined to make one part G, which was simplified by asking for only a yes/no response to the following question: "Does your library provide the following?" All but three items from part G1 were dropped and one item was added, as follows:

Dropped items
- documents digitized by the library staff;

- library reference service by e-mail or on the Web; and

- technology to assist patrons with disabilities (e.g., TDD, specially equipped workstations).

New item
- electronic theses and dissertations.

The 2004 ALS added a set of questions on "information literacy" to the survey instrument, including the following:

- Is the library collection entirely electronic?

- Were electronic reference sources and aggregation services added?

- Were electronic reference sources and aggregation services held?

- Does your library have a definition of information literacy or of an information-literate student?

- Has your library incorporated information literacy into the institution's mission?

- Has your library incorporated information literacy into the institution's strategic plan?

- Does your library have an institution-wide committee to implement the strategic plan for information literacy?

The 2006 ALS added another question on information literacy to the survey instrument:

- Does the strategic plan formally recognize the library's role in information literacy instruction?

In 2008, the eligibility questions were revised as follows:

- The financial support question was deleted.

- The first sentence was updated.

- Questions b and c were revised to add "paid, trained staff"

The 2010 survey instrument underwent the following changes:

Eligibility Questions: A new question was added – Does your total library expenditures exceed $10,000?

Library Services, FY 2010 section:

- Change in instructions for general circulation transactions from "Report the number of items lent from the general collection. Include both initial transactions and renewals." to "Report the number of items lent from the general collection (all formats). Include both initial transactions and renewals."

- New section was added – Information services to individuals. The new questions were as follows: In person reference, virtual reference,

total reference, in person consultations, virtual consultations and total consultations.

Library Services, Typical week, FY 2010 section:

- Reference transaction in a typical week was changed to total information service to individuals (a yearly figure now reported in Library services, FY 2010 sections).

Virtual Reference section: A new set of questions was added:

- Does your library support virtual reference services? If no, select “N” and skip 901 thru 904.

- If yes, does your library utilize any of the following and does it collect usage statistics from any of the virtual reference utilities?

    - E-mail reference

    - Chat reference, commercial service

    - Chat reference, instant messaging application

    - Short message service (SMS) or text messaging

Future Plans

At this time, NCES plans to continue conducting ALS biennially.

5. DATA QUALITY AND COMPARABILITY

NCES makes every effort to achieve high data quality. Through a web collection that includes built-in edit checks, it hopes to improve the quality of ALS data. Users are cautioned about limitations in the analysis of ALS data by state or by level and control of institution. Since nonresponse varies by state, the reliability of state estimates and comparisons is affected. Special caution should be exercised when using data where the nonresponse rate is 30 percent or greater. See below for more information on the types of errors that affect data quality and comparability.

Sampling Error

Because ALS is a universe survey, there is no sampling error.

Nonsampling Error

Coverage error. A comprehensive evaluation of the coverage of ALS found that the quality of institutional coverage was excellent (a coverage gap of only 1 to 3 percent) when compared to other institutional listings directly related to the academic libraries industry; however, questions remain as to whether the data collected by ALS fully account for branch data associated with parent institution resources. (See *Coverage Evaluation of the Academic Library Survey* [Marston 1999]) A second problem is that the ALS data for some parent colleges or universities may not contain statistics for their professional schools.

Nonresponse error.

Unit nonresponse. The overall unit response rate for the 2000 ALS was 87.4 percent. Four-year institutions had a response rate of 88.5 percent (from a low of 85.5 percent at the bachelor’s level to a high of 91.0 percent at the doctor’s level), while less-than-4-year institutions had a response rate of 85.8 percent. The response rate was 93.3 percent for public institutions and 82.8 percent for private institutions.

For the 2002 ALS, the overall unit response rate was 88.6 percent. The response rate for all 4-year institutions was 89.8 percent (85.9 percent at the bachelor’s level, 89.6 percent at the master’s level, and 91.0 percent at the doctor’s level). Less-than-4-year institutions had a response rate of 86.6. Public institutions had a response rate of 93.4 percent, while private institutions had a response rate of 84.6 percent.

The overall unit response rate for the 2004 ALS was 87.0 percent. The aggregate response rate for 4-year institutions was 88.8 percent (ranging from 86.5 percent at the bachelor’s level to 91.3 percent at the doctor’s level). Less-than-4-year institutions had a slightly lower response rate (84.3 percent). The response rate was 92.2 percent for public institutions and 83.0 percent for private institutions.

The overall unit response rate for the 2006 ALS was 88.8 percent. The response rate for all 4-year institutions was 90.0 percent (89.2 percent at the bachelor’s level, 89.6 percent at the master’s level, and 91.7 percent at the doctor’s level). The overall response rate for less-than-4-year institutions was 86.7 percent (93.0 percent for public institutions and 85.6 percent for private institutions).

The overall unit response rate for the 2008 ALS was 86.7 percent. The response rate for all 4-year institutions was 87.1 percent (80.0 percent at the bachelor’s level, 91.0 percent at the master’s level, and 89.3 percent at the doctor’s level). The overall response

rate for less-than-4-year institutions was 86.1 percent (95.4 percent for public institutions and 80.6 percent for private institutions).

Item nonresponse. For the 2000 ALS, overall item response rates ranged from 68.6 to 86.9 percent. Out of 102 questions, 86 had response rates at or above 80 percent. Five items had response rates below 75 percent: one in the area of library staff (74.1 percent), two in the area of library operating expenditures (74.7 percent and 73.4 percent), and two in the area of library collections (73.2 percent and 73.8 percent).

Overall item response rates in the 2002 ALS ranged from to 57.4 to 100.0 percent. Of the 57 questions, 52 had response rates at or above 80 percent. Three items had response rates below 75 percent: two in the area of library services (73.2 percent and 72.7 percent) and one in the area of library collections (57.4 percent).

In the 2004 ALS, overall item response rates ranged from 73.4 to 86.7 percent. Of the 63 questions, 58 had response rates at or above 80 percent. Only two items has response rates below 75 percent, both in the area of library collections (73.4 percent and 74.3 percent).

Overall item response rates in 2006 ranged from 78.9 to 88.8 percent. Of the 60 questions, 59 had a response rate at or above 80 percent. No item had a response rate below 75 percent.

Overall item response rates in 2008 ranged from 71.8 to 86.3 percent. Three items had a response rate below 75 percent.

Measurement error. No information is available.

6. CONTACT INFORMATION

For content information on ALS, contact:

Tai Phan
Phone: (202) 502-7301
E-mail: tai.phan@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Cahalan, M.W., and Justh, N.M. (2000). *Academic Libraries: 1996* (NCES 2000-326). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Cahalan, M.W., and Justh, N.M. (2001). *Academic Libraries: 1998* (NCES 2001-341). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Carey, N., and Justh, N.M. (2003). *Academic Libraries: 2000* (NCES 2004-317). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Holton, B., and George, A. (2007). *Data File and Documentation for the Academic Libraries Survey (ALS): 1996 (Public Use)* (NCES 2008-318). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Holton, B., Hardesty, L., and O'Shea, P. (2008). *Academic Libraries: 2006* (NCES 2008-337). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Holton, B., Hardesty, L., and O'Shea, P. (2008). *Documentation for the 2006 Academic Libraries Survey Restricted-Use Data File* (NCES 2008-350). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Holton, B., Hardesty, L., and O'Shea, P. (2008). *Documentation for the Academic Library Survey (ALS) Public Use Data File: 2006* (NCES 2008-348). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Holton, B., Vaden, K., and O'Shea, P. (2001). *Academic Libraries: 2004* (NCES 2007-301). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Phan, T., Hardesty, L., Sheckells, C., and Davis, D. (2009). *Academic Libraries: 2008* (NCES 2010–348). National Center for Education Statistics,

Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Phan, T., Hardesty, L., and Sheckells, C. (2009). *Documentation for the Academic Library Survey (ALS) Data File: Fiscal Year 2008 (Public Use)* (NCES 2010–311). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Schmitt, C.M. (2005). *Documentation for the Academic Library Survey (ALS) Data File: Fiscal Year 2002 (Public Use)* (NCES 2006-308). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Schmitt, C.M. (2006). *Documentation for the Academic Library Survey (ALS) Data File: Fiscal Year 2000 (Public Use)* (NCES 2006-342). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Schmitt, C.M., O'Shea, P., and Vaden, K. (2007). *Documentation for the Academic Library Survey (ALS) Data File: Fiscal Year 2004 (Public Use)* (NCES 2007-343). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Cahalan, M., Mansfield, W., and Justh, N. (2001). *Status of Academic Libraries in the United States: Results From the 1996 Academic Library Survey With Historical Comparisons* (NCES 2001-301). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Marston, C.C. (1999). *Coverage Evaluation of the Academic Library Survey* (NCES 1999-330). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 12: Integrated Postsecondary Education Data System (IPEDS)

1. OVERVIEW

The Integrated Postsecondary Education Data System (IPEDS) is NCES's core postsecondary education data collection program, designed to help NCES meet its mandate to report full and complete statistics on the condition of postsecondary education in the United States. The IPEDS collects institution-level data from providers of postsecondary education in the United States (the 50 states and the District of Columbia) and other jurisdictions, such as the U.S. Virgin Islands. The IPEDS is a single, comprehensive system that is built around a series of interrelated surveys designed to collect institution-level data in such areas as enrollment, program completions, graduation rates, student financial aid, tuition and fees, faculty, staff, and finances.

Beginning in 1993, the IPEDS survey completion became mandatory for all postsecondary institutions with a Program Participation Agreement (PPA) with the Office of Postsecondary Education (OPE), U.S. Department of Education—that is, institutions that participate in or are eligible to participate in any federal student financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended (20 USC 1094[a] [17]). For institutions not eligible to participate in Title IV programs, participation in the IPEDS is voluntary. Prior to 1993, only national-level estimates from a sample of institutions are available for private less-than-2-year institutions.

In 1998, due to several externally mandated changes and additions to the IPEDS, changes in technology for data collection and dissemination, changes in postsecondary education issues, and new expectations for the IPEDS, a redesign task force was charged with recommending changes for the system. The primary recommendation was that the IPEDS switch from paper forms to a solely web-based reporting system. The IPEDS program was completely redesigned for the 2000–01 survey year, and the data collection was converted from a paper-based to a fully web-based system. The web-based survey instruments offered many features to improve the quality and timeliness of the data. Currently, the IPEDS remains an annual survey, with data collection occurring three times per year: in fall, winter, and spring. The next data collection is scheduled for the fall of 2010.

It was in 1986 that the IPEDS replaced the Higher Education General Information Survey (HEGIS). HEGIS collected data from 1966 to 1986 from a more limited universe of approximately 3,400 institutions accredited at the college level by an association recognized by the Secretary of the U.S. Department of Education. The transition to the IPEDS program expanded the universe to include *all* institution whose primary purpose is the provision of postsecondary education. The system currently includes about 7,000 Title IV institutions and 200 non-Title IV institutions—including many schools not accredited at the college level but with vocational/occupational accreditation. Note that the U.S. Department of Education's Office for Civil Rights (OCR) has collaborated with NCES since 1976 on the collection of data from postsecondary institutions through compliance reports

SURVEY OF THE UNIVERSE OF POSTSECONDARY INSTITUTIONS

IPEDS collects data annually or biennially through these major components:

- Institutional Characteristics
- Completions
- 12-Month Enrollment
- Fall Enrollment
- Graduation Rates
- Student Financial Aid
- Human Resources
- Finance

mandated pursuant to Title VI of the Civil Rights Act of 1964, first through HEGIS and then through the IPEDS.

Purpose

To collect institution-level data from all Title IV providers of postsecondary education—universities and colleges, as well as institutions offering technical and vocational education beyond the high school level.

Components

The IPEDS program consists of several components that obtain information on who provides postsecondary education (institutions), who participates in it and completes it (students), what programs are offered, what programs are completed, and the human and financial resources involved in the provision of institution-based postsecondary education. To avoid duplicative reporting and thus enhance the analytic potential of the database, the various IPEDS data elements and component surveys are interrelated. Survey components are tailored to the institution using institutional characteristics. In general, the most extensive data are collected from postsecondary institutions granting baccalaureate and higher degrees; less extensive data are requested from other types of institutions. This feature accommodates the varied operating characteristics, program offerings, and reporting capabilities of postsecondary institutions while yielding comparable statistics for all institutions.

The IPEDS program currently collects information from approximately 7,200 postsecondary institutions using a combination of survey components. Participation in the IPEDS is a requirement for institutions that participate in Title IV federal student financial aid programs, such as Pell Grants or Stafford loans. Title IV institutions include traditional colleges and universities, 2-year institutions, and for-profit degree- and non-degree-granting institutions (such as schools of cosmetology), among others. Because of the requirements for participation in Title IV federal financial aid programs, the IPEDS focuses on the institutions designated as Title IV participants (about 7,200 institutions). Institutions that do not participate in Title IV programs may participate in the IPEDS data collection on a voluntary basis.

The IPEDS collects data three times per year—in fall, winter, and spring—using the following instruments. The Institutional Characteristics, Completions, and 12-month Enrollment surveys are administered in the fall. The Human Resources component (consisting of the Employees by Assigned Position, Salaries, and Fall Staff sections), is collected in the winter, and the Fall Enrollment and Finance surveys, are administered in the winter. (Institutions can also elect to submit fall enrollment and finance data in the spring.) The Student Financial Aid and Finance components are administered in the spring.

Each of these instruments (or components) is described below; the abbreviation for the survey component is provided after the component name.

Institutional Characteristics (IC). The core of the IPEDS program is the annual Institutional Characteristics component collected each fall—intended for completion by all currently operating postsecondary institutions in the United States and its other jurisdictions. As the control file for the entire IPEDS program, IC constitutes the sampling frame for all other NCES surveys of postsecondary institutions. It also helps determine the specific IPEDS screens that are shown to each institution (as it used to determine the specific survey forms that were mailed to each institution). This component collects basic data on each institution, such as identification; educational offerings; control or affiliation; tuition; room and board charges; admission requirements; levels of degrees and awards; estimated fall enrollment; and student services. These data are necessary to sort and analyze not only the IC data file, but also all the other IPEDS compontent data files. The IC Survey incorporates many data elements required by state career information delivery systems, thereby reducing or eliminating the need for these organizations to conduct their own surveys.

IC data are collected for the academic year, which generally extends from September of one calendar year to June of the following year. Specific data elements currently collected for each institution include the institution name, address, telephone number, web address, control or affiliation, calendar system, levels of degrees and awards offered, types of programs, application and admissions information, and student services offered. The IC component also collects information on tuition and required fees, room and board charges, books and supplies, and other expenses for release on NCES's College Navigator website (http://nces.ed.gov/collegenavigator/). The College Navigator is designed to help college students, prospective students, and their parents understand the differences among colleges and how much it costs to attend college, as well as offer information on student financial aid, programs and services offered, enrollments and graduation rates, and accreditation, among other things.

Completions (C). The Completions component collects data annually each fall on recognized degree completions in postsecondary education programs by

level (associate's, bachelor's, master's, doctor's, and professional) and on other formal awards by length of program. These data are collected by race/ethnicity and gender of recipient and by fields of study, which are identified by 6-digit Classification of Instructional Programs (CIP) codes from the NCES publication *Classification of Instructional Programs* (http://nces.ed.gov/ipeds/cipcode/). Completions data on multiple majors are collected by 6-digit CIP code, award level, race/ethnicity, and gender from those schools that award degrees with multiple majors.

OCR has provided support to collect Completions data since 1976.

12-Month Enrollment (E12). This annual component in the fall collection collects 12-month enrollment data for award levels ranging from postsecondary certificates of less than 1 year to doctoral degrees. The component collects data on unduplicated headcounts and instructional activity (contact or credit hours). A standardized, 12-month full-time equivalent (FTE) enrollment is computed based on instructional activity, and institutions may report an alternate FTE as well. The headcount data collected include demographic information on race/ethnicity and sex. Data are collected for a 12-month reporting period in the previous year; institutions must indicate the 12-month period for which they are reporting—either July 1 through June 30, or September 1 through August 31.

Fall Enrollment (EF). This spring collection component collects data annually on the number of full- and part-time students enrolled in postsecondary institutions in the United States and its other jurisdictions, by level (undergraduate, graduate) and by race/ethnicity and gender of student.

Institutions report on students enrolled in courses creditable toward a degree or other formal award; students enrolled in courses that are part of a vocational or occupational program, including those enrolled in off-campus centers; and high school students taking regular college courses for credit. An item that asks for the total number of undergraduates in the entering class (including first-time, transfer, and nondegree students) was added in 2001. Full- and part-time fall-to-fall retention rates for first-time, degree/certificate-seeking students are also collected.

Age distributions are collected in odd-numbered years by student level. Data on the state of residence of first-time freshmen (first-time, first-year students) and the number of students who graduated from high school in the past 12 months are collected in even-numbered years (replacing an earlier survey on Residence of First-Time Students). Four-year institutions are also required, in even-numbered years, to complete enrollment data by level, race/ethnicity, and gender for nine selected fields of study—Education, Engineering, Law, Biological Sciences/Life Sciences, Mathematics, Physical Sciences, Dentistry, Medicine, and Business Management and Administrative Services. The specified fields and their codes are taken directly from *Classification of Instructional Programs*.

OCR has supported the collection of these data since 1976.

Fall Enrollment in Occupationally Specific Programs (EP). This component was incorporated into the IPEDS program in response to the Carl Perkins vocational education legislation. Conducted biennially in odd-numbered years, this survey collected fall enrollment data on students enrolled in occupationally specific programs at the subbaccalaureate level, by race/ethnicity and gender of student and by fields of study (identified by 6-digit CIP codes). Starting in 1995, total unduplicated counts of students enrolled in these programs were also requested. This survey was discontinued as of the 1999–2000 data collection.

Graduation Rate Survey (GRS). This annual spring collectin component was added in 1997 to help institutions satisfy the requirements of the Student Right-to-Know Act of 1990. For the 1997–98 GRS, 4-year institutions reported on a 1991 cohort, and less than 4-year institutions reported on a 1994 cohort.

Institutions provide data on their initial cohort of full-time, first-time, degree/certificate-seeking undergraduate students; on the number of those students completing within 150 percent of the normal time; and on the number of students who transferred to other institutions. Four-year institutions report separately on their bachelor's degree-seeking students. Data are reported by race/ethnicity and gender. These data allow institutions to disclose and/or report information on the completion or graduation rates and transfer-out rates of their students. Worksheets automatically calculate rates within the web system.

A supplemental form is used to collect data on students who completed a long program within 150% of normal time, e.g., a 5-year bachelor's degree program or 3-year associate's degree program.

One hundred percent graduation rates data are also collected. Four-year bachelor's rates have been reported by 4-year institutions since 1997, and 100% rates have been reported by less than 4-year institutions since 2008-09.

200% Graduation Rates (GR200). This survey component was added to the spring collection in 2009–10. It is separate from the regular GRS component so not to confuse the two different cohorts that are being reported on. The *GR200* asks institutions to report additional data on cohort students so that 200% graduation rates can be calculated. Graduation rates at 200 percent of normal time are calculated for full-time, first-time bachelor degree-seeking students at 4-year institutions, and for all full-time, first-time degree/certificate-seeking undergraduate students at less than 4-year institutions.

Student Financial Aid (SFA). This spring collection component collects student financial aid data on several different student populations: undergraduate students; a cohort of full-time, first-time, degree/certificate-seeking undergraduate students; and two subpopulations of that cohort. The financial aid data collected on the subpopulations is used to calculate the institution's average net price of attendance, and average net price of attendance by income category. Data are collected for the previous aid year. Number of students receiving aid and total amount of aid received are collected for different aid types; average amount of aid received by type of aid and percent of students receiving aid by type of aid are calculated. For undergraduates, total grant or scholarship aid, Pell grants, and federal loans are the aid types. For the cohort, aid types are federal grants (Pell grants and other federal grants), state/local government grants or scholarships, institutional grants or scholarships, and loans to students (total loans, Federal loans, other loans).

This component began with a pilot test in 1999 and collected both pricing and student financial aid data. The pricing items are now part of the Institutional Characteristics Survey, conducted annually in the fall; the SFA component is part of the annual spring data collection.

Human Resources (HR). The Human Resources component, collected in the water consists of three sections: Employees by Assigned Position, Fall Staff, and Salaries. These three sections (see below) were previously separate components, but were merged into the single HR component beginning with the 2005–06 survey year in order to simplify reporting and ensure data consistency and accuracy.

Employees by Assigned Position (EAP). Beginning with the winter 2001-02 collection, a new annual survey, Employees by Assigned Position, proposed by the National Postsecondary Education Cooperative focus group on faculty and staff, was instituted. This survey was optional in the first year, but became mandatory in 2002–03. The EAP section categorizes all staff on the institution's payroll as of November 1 of the collection year by full- and part-time status; by function or occupational category; and by faculty status and tenure status (if applicable). Institutions with medical schools are required to report their medical school data separately. The medical school pages of EAP are applicable to institutions with M.D. and/or D.O. programs only. Employees who are in health disciplines that are not considered part of the medical school are reported in the nonmedical school part of EAP.

Fall Staff (S). This survey is conducted biennially in odd-numbered years and collects data on the numbers of full- and part-time institutional staff and includes demographic information on race/ethnicity and gender. (During even-numbered years, reporting Fall Staff data is optional.) Specific data elements include number of full-time staff by contract length and salary class intervals; number of other persons employed full time by primary occupational activity and salary class intervals; part-time employees by primary occupational activity; tenure of full-time faculty by academic rank; and new hires by primary occupational activity.

Between 1987 and 1991, the Fall Staff data were collected in cooperation with the U.S. Equal Employment Opportunity Commission (EEOC). From 1976 through 1991, EEOC collected data on staff through its biennial Higher Education Staff Information (EEO-6) report from all postsecondary institutions within its mandate—that is, institutions that had 15 or more full-time employees. Through the IPEDS program, NCES collected data from all other postsecondary institutions, including all 2- and 4-year higher education institutions with fewer than 15 full-time employees and a sample of less-than-2-year schools. The 1987–91 IPEDS Fall Staff data files contain combined data from the EEO-6 and the IPEDS staff surveys. Beginning in 1993, all schools formerly surveyed by EEOC were surveyed through the IPEDS Fall Staff Survey.

OCR began supporting the collection of these data in 1993.

Salaries (SA) (formerly Salaries, Tenure, and Fringe Benefits of Full-Time Instructional Faculty). The primary purpose of this section is to collect data on the salaries, tenure, and fringe benefits of full-time instructional staff (referred to as instructional faculty prior to the 2005–06 survey year) by contract length, gender, and academic rank. Institutions are excluded from completing the Salaries section if all of their instructional staff (1) are employed on a part-time

basis, (2) were military personnel, (3) contributed their services (e.g., members of a religious order), or (4) teach preclinical or clinical medicine.

Data are collected on total salary outlays; total number of full-time instructional staff paid these outlays; and number of staff members with tenure, on tenure track, and not on tenure track. These data are collected by rank (professor, associate professor, assistant professor, instructor, lecturer, no academic rank) for men and women on 9- or 10-month and 11- or 12-month contracts or teaching periods. Fringe benefits are collected for instructional staff on 9/10-month and 11/12-month contracts or teaching periods. Specific data elements are included for retirement, tuition, housing, medical/dental plans, group life insurance, unemployment and worker's compensation, social security taxes, fringe benefit expenditures (in whole dollars), and the number of full-time staff covered, by length of contract contract or teaching period.

This Salaries data collection was changed from a biennial to an annual collection in 1990, and data were not collected in 2000.

Finance (F). This component, collected in the spring, collects summary data on each institution's financial status in the applicable fiscal year. The Finance component has different versions of the form based mainly on control of the institution: public, private not-for-profit, and private for-profit. The primary purpose of this annual component is to collect data to describe the financial condition of postsecondary education in the nation; to enable changes in postsecondary education finance to be monitored; and to promote research involving institutional financial resources and expenditures.

For public institutions that use Governmental Accounting Standards Board (GASB) reporting standards to prepare their financial statements, data are collected on statement of net assets, plant, property, and equipment, revenues and other additions, expenses and other deductions, summary of changes in net assets, scholarships and fellowships, and endowment assets. Additionally, certain data are collected for the U.S. Bureau of the Census, including revenue data, expenditure data, and debts and assets

Private not-for-profit institutions and public institutions that use Financial Accounting Standards Board (FASB) reporting standards to prepare their financial statements report data on their statement of financial position, summary of changes in net assets, student grants, revenues and investment return, expenses by functional and natural classification, and endowment assets. A shortened version of the not-for-profit form has been developed for private for-profit institutions, and data are collected on balance sheet information, summary of changes in equity, student grants, revenues and investment return, and expenses by function.

A 2-year phase-in period began with FY 2008 reporting to implement additional changes to better align the finance reporting of public and private institutions. For FY 2010 reporting, all public and not-for-profit institutions used the new aligned form.

Academic Libraries. First administered in 1966, the Academic Libraries Survey was designed to provide concise information on library resources, services, and expenditures for the entire population of academic libraries in the United States. In 1988, the Academic Libraries Survey became a part of the IPEDS program and was conducted biennially in even-numbered years. From 1966 to 1988, the Academic Libraries Survey was conducted on a 3-year cycle. As of September 2000, this survey ceased to be a part of the IPEDS. (See chapter 11 for a full description of the Academic Libraries Survey.)

Consolidated Form (CN and CN-F). When paper survey forms were used, a Consolidated Form was used to collect the IPEDS data from the institutions that did not complete the full package of the IPEDS components described above—that is, accredited institutions granting only certificates at the subbaccalaureate level. The Consolidated Form consisted of four or five parts designed to collect, on the same schedule as the regular IPEDS components, minimal data on enrollment (including occupationally specific programs) and completions by race/ethnicity and gender, as well as data on finance, fall staff, and academic libraries. As of 1996, the Finance part of the Consolidated Form was moved to a separate form (CN-F). The purpose and use of the Consolidated Form was the same as for the full package of surveys: to allow national data on all accredited institutions to be presented and analyzed. The Consolidated Form is no longer needed, since the web-based data collection system, implemented in the 2000–01 survey year, automatically tailors data items for institutions based on selected characteristics and screening questions.

Periodicity

The IPEDS program replaced the HEGIS program in 1986. The IPEDS data were collected on paper forms between 1986 and 1999. Since the implementation of the web-based collection of the IPEDS data in 2000, most components are completed by institutions on an annual basis. However, the components schedules vary slightly. The Institutional Characteristics, Fall

Enrollment, 12-month Enrollment, Completions, Graduation Rate, Student Financial Aid, and Finance components are conducted annually. The Salaries Survey is also annual, except for the 2000–01 collection. Within the Fall Enrollment component, the Age and Residence sections alternate, but are available in the off years for those institutions wishing to submit the data; the collection of enrollment by program is done only in even-numbered years. The Human Resouces component is also annual; the Employees by Assigned Position section and Salaries section are collected yearly (Salaries was not collected in 2000-01), and the Fall Staff section continues to be conducted on a biennial basis in odd-numbered years (but is available in even-numbered years if institutions wish to submit those data).

2. USES OF DATA

The IPEDS surveys provide a wealth of national-, state-, and institution-level data for analyzing the condition of postsecondary education institutions. For example, the data can be used (with the earlier HEGIS data) to describe long-term trends in higher education. NCES uses the IPEDS data in annual reports to Congress on the condition of postsecondary education, statistical digests, profiles of higher education in the states, and other publications. In addition, many requests for information based on the IPEDS surveys are received each year from Congress, federal agencies and officials, state agencies and officials, education associations, individual institutions, the media, and the general public. Federal program staff use the IPEDS data to address various policy issues. State policymakers use the IPEDS data for planning purposes and comparative analysis. Institutional staff use the data for peer analysis.

The IPEDS data respond to a wide range of specific educational issues and public concerns. Policymakers and researchers can analyze the types and numbers of postsecondary institutions; the number of students, graduates, first-time freshmen, and graduate and professional students by race/ethnicity and gender; the status of postsecondary vocational education programs; the number of individuals trained in certain occupational and vocational fields by race/ethnicity, gender, and level; the resources generated by postsecondary institutions; patterns of expenditures and revenues of institutions; changes in tuition and fees charged and student financial aid recieved; completions by type of program, level of award, race/ethnicity, and gender; faculty composition and salaries; and many other topics of interest.

The IPEDS universe also provides the institutional sampling frame used in all NCES postsecondary surveys, such as the National Postsecondary Student Aid Study and the National Study of Postsecondary Faculty. Each of these surveys uses the IPEDS institutional universe for its first-stage sample and relies on the IPEDS results on enrollment, completions, or staff to weight its second-stage sample.

OCR supports the collection of the IPEDS enrollment, completions, and fall staff data, and uses these data to produce reports.

3. KEY CONCEPTS

Key Terms

Described below are several key concepts relevant to the IPEDS program. For additional terms, refer to the *IPEDS Glossary* at http://nces.ed.gov/ipeds/glossary.

Postsecondary Education. The provision of a formal instructional program whose curriculum is designed primarily for students who are beyond the compulsory age for high school. This includes programs whose purpose is academic, vocational, or continuing professional education , and excludes avocational and adult basic education programs.

Postsecondary Education Institution. An institution which has as its sole purpose or one of its primary missions, the provision of postsecondary education.

Institution of Higher Education (IHE). Prior to 1996, an IHE was defined as an institution accredited at the *college* level by an accrediting agency or association recognized by the Secretary of the U.S. Department of Education—and indicated as such in the database by the presence of a Federal Interagency Committee on Education (FICE) code. IHEs were legally authorized to offer at least a 1-year program of study creditable toward a degree.

Degree-Granting Institution. Any institution offering an associate's, bachelor's, master's, doctor's, or first-professional degree. Institutions that grant only certificates or awards of any length (less than 2 years, or 2 years or more) are categorized as nondegree-granting institutions.

Branch Institution. A campus or site of an educational institution that is not temporary, that is located in a community beyond a reasonable commuting distance from its parent institution, and that offers full programs of study (not just courses). This last criterion is the

most important. It means that at least one degree or award program can be completed entirely at the site without requiring any attendance at the main campus or any other institution within the system.

OPEID Code. An 8-digit identification code developed by the U.S. Department of Education's Office of Postsecondary Education (OPE) for the Postsecondary Education Participants System (PEPS). The presence of a valid OPEID in the database indicates that the school has a PPA with the Department of Education and is currently eligible to participate in Title IV federal financial aid programs (e.g., Pell grants, Stafford loans, college work-study). The first 6 digits of the OPEID are the old FICE code and identify the institution. The last 2 digits identify the various campuses or additional locations. For the main campus, the last 2 digits will always be "00." If the last 2 digits are numeric (e.g., 01, 02, 03), the institution is a branch campus or other location of an eligible main campus and is listed separately in PEPS. If the last 2 digits of the OPEID are of the form A1, A2, etc., the entity is separately identified in the IPEDS for reporting purposes.

Occupationally Specific Program. An instructional program below the bachelor's level that is designed to prepare individuals with the entry-level skills and training required for employment in a specific trade, occupation, or profession related to the field of study.

CIP Code. A 6-digit code, in the form xx.xxxx, that identifies instructional program specialties within educational institutions. The codes are from the NCES publication *Classification of Instructional Programs* (http://nces.ed.gov/ipeds/cipcode/).

4. SURVEY DESIGN

Target Population

All institutions (in the 50 states, the District of Columbia, and other jurisdictions) whose purpose is the provision of postsecondary education may participate in IPEDS, but the majority of institutions represented are those that are eligible to participate in Title IV federal student financial aid programs. The IPEDS universe includes institutions and branch campuses that offer a full program of study (not just courses); freestanding medical schools, as well as schools of nursing, schools of radiology, etc., within hospitals; and schools offering occupational and vocational training with the intent of preparing students for work (e.g., a modeling school that trains for professional modeling, but not a charm school).

The IPEDS universe of postsecondary institutions does *not* include institutions that are not open to the general public (training sites at prisons, military installations, corporations); hospitals that offer only internships or residency programs or that offer only training as part of a medical school program at an institution of higher education; organizational entities providing only noncredit continuing education; schools whose only purpose is to prepare students to take a particular test, such as the CPA or bar exams; and branch campuses of U.S. institutions in foreign countries. Relevant data from such locations or training sites are to be incorporated into the data reported by the main campus or any other institution or branch campus in the system that is most appropriate. Prior to 2010-11, Title IV institutions that are not primarily postsecondary (e.g., secondary technical schools with a small postsecondary component) reported to IPEDS voluntarily; starting in 2010-11 their participation is required.

Eligibility for Title IV federal financial aid, while not a requirement for inclusion in the universe, defines a major subset of all postsecondary institutions. Prior to 1996, aid-eligible institutions were self-identified as IHEs or were identified as aid-eligible from responses to items in the Institutional Characteristics Survey. Since 1996, the subset of aid-eligible institutions has been validated by matching the IPEDS universe with the PEPS file maintained by OPE. OPE grants eligibility to institutions to participate in Title IV federal financial aid programs.

In establishing the PEPS file, the U.S. Department of Education discontinued its tradition of distinguishing institutions accredited at the college level from institutions accredited at the occupational/vocational level. Therefore, it is no longer possible for NCES to maintain a subset of accredited institutions at the college level (IHEs). Beginning with the 1997 IPEDS mailout and in the 1996 and subsequent data files, institutions have been classified by whether or not they are eligible to participate in Title IV financial aid programs and whether or not they grant degrees (as opposed to awarding only certificates).

Sample Design

Prior to 1993, data were collected from a representative sample of about 15 percent of the universe of private, for-profit, less-than-2-year institutions. However, the Higher Education Act of 1992 mandated the completion of the IPEDS surveys for all institutions that participate in or are applicants for participation in any federal student financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended. Thus, beginning with the 1993

IPEDS mailout, NCES surveys in detail *all* postsecondary institutions meeting this mandate.

Data Collection and Processing

The U.S. Bureau of the Census served as the data collection agent for the IPEDS surveys from 1990 through the 1999–2000 survey. Survey forms were either submitted directly to the Census Bureau by the institutions or through a central or state coordinating office. The web-based data collection system was implemented with the 2000–01 survey, with different contractors developing the website and managing the collection process.

The IPEDS institution-level data collection allows for aggregation of results at various levels and permits significant controls on data quality through editing. Attempts are made to minimize institutional respondent burden by coordinating data collection with the states and with other offices and agencies that regularly collect data from institutions.

Reference dates. Data for the IPEDS component surveys are collected for a particular academic year, 12-month period, or fiscal year, as follows:

- The Institutional Characteristics component collects data for the entire current academic year, generally starting in September, or with the fall term, if there is one. In the case of schools operating on a 12-month calendar, the reference period runs from the current September through August.

- The Completions component collects data for an entire 12-month period, which is defined as July 1 through June 30; in some instances, start dates may vary slightly by institution.

- The 12-month Enrollment component collects data for a 12-month reporting period in the previous year; institutions must indicate the 12-month period for which they are reporting—either July 1 through June 30, or September 1 through August 31.

- The Fall Enrollment component (and previously the Fall Enrollment in Occupationally Specific Programs component) collects data for a single point in time during the fall term, usually recorded as of the institution's official fall reporting date or October 15. Institutions that operate on a continuous basis report their fall enrollment based on the time period between August 1 and October 31. If there is no fall term or class activity, institutions are asked to report zero enrollment.

- For the Graduation Rate component, institutions report on the status of students in their cohort (either a fall cohort or a full-year cohort) as of August 31.

- The Student Financial Aid component collects data for the prior aid year. Institutions reporting on a fall cohort report aid for the prior academic year; institutions reporting on a full-year cohort report aid for the prior 12-month period.

- The Employees by Assigned Position and Fall Staff sections of the Human Resources component collects data on staff on the institution's payroll as of November 1 of the current academic year. Additionally, the Fall Staff section collects data on new hires from July 1 through October 31 of the survey year. Prior to the 2001 collection, institutions reported as of October 1. Salaries and fringe benefits data collected in the Salaries section reflect the full academic year.

- The Finance component collects data for the institution's most recent fiscal year ending before October 1. Thus, data collected in spring 2010 (part of the 2009-10 data collection cycle) pertain to the fiscal year just ended, FY 2009.

Data collection. Since institutions are the primary unit of data collection, institutional units must be defined as consistently as possible. The IPEDS program does not request separate reports from more than one component within an individual institution; however, separate branch campuses are asked to report as individual units. Following the HEGIS model, the IPEDS program is intended to collect data from each institution in a multi-institutional system and each separate branch in a multi-campus system.

Schools targeted as "possible adds" are identified from many sources, including a review of the PEPS data file from OPE, and information received from the institutions themselves. Institutions are added to the universe if they respond that they provide postsecondary education as defined in the survey. Unlike in past years (prior to 2000), these institutions submit all survey components in their first year in IPEDS.

Institutions found to be closed or out-of-scope during data collection are deleted from the IPEDS universe. These deletions result from formal notification from

the Postsecondary Education Participation System, the institution, or the IPEDS state coordinators. Included in the deletions are (1) duplicates of other institutions on the file; (2) institutions that closed or merged with another institution and, thus, are no longer legitimate institutions or branches; (3) institutions that no longer offer postsecondary programs; and (4) schools that do not conform to the IPEDS definition of an institution or branch. The final IPEDS universe is also adjusted to reflect institutions that have changed from one sector to another.

Institutions receive letters or emails in August containing UserIDs and passwords for the web-based data collection system, and instructions for registering their keyholder. The keyholder is responsible for entering and locking the institution’s data by each collection close date. Follow-up is done by email and telephone, and is conducted either directly with the keyholder or with the institution’s chief executive officer (if there is no registered keyholder). State IPEDS coordinators also conduct follow-up.

To ease respondent burden, the Institutional Characteristics web screens include previously reported data, and survey respondents are instructed to update the previous data, if necessary, and to provide current information for items such as tuition and required fees, and room and board charges. (In earlier years, IC forms were preprinted with prior-year survey responses for those items that generally were not expected to change from year to year.) Screens for other IPEDS components contain selected information from previous reporting (such as CIP codes and program titles in the screens for the Completions and Enrollment components, and cohort for Graduation Rates). Prior year values are preloaded on screens for reference and to edit against, and values are brought forward from one section to another where they must match. Totals, differences, and rates are calculated by the data collection system. Institutions may choose to key enter their data into the system, or to upload a file in a fixed, key value, or, more recently, XML format.

State and system IPEDS coordinators play a large role in the submission and review of IPEDS data. In many states, the IPEDS institutional data are provided by the state higher education agency from data collected in state surveys. Coordinators may choose the sectors and institutions they wish to monitor (e.g., they can identify just 4-year schools or specify particular institutions); they can also choose to view the data only, or actually review, approve, and “lock” the data. Alternatively, state agencies may extract data from the IPEDS rather than conduct their own surveys.

Prior to web-based data collection, mailouts of survey forms generally took place in July of the survey year. Due dates varied by component. Extensive follow-up for survey nonresponse was conducted during the 6 months following each component’s due date. Initially, reminder letters were mailed, encouraging nonresponding institutions to complete and return their forms. Subsequently, the Postsecondary Education Telephone System (PETS) was used to collect critical data by telephone from representatives of institutions for which the IPEDS state coordinators were not responsible for follow-up.

Institutions reported the IPEDS data by mail (on paper forms or diskettes), by fax, or electronically through the Internet. Two methods were available: the first method involved a predetermined ASCII record layout, available for all surveys, except Institutional Characteristics. For the Fall Enrollment and Graduation Rate surveys, a second method was available that used downloadable software for data entry as well as preliminary editing of the data before transmission to the Census Bureau.

The current IPEDS universe includes approximately 7,200 postsecondary institutions and 84 administrative units (central and system offices).

Editing. Edit checks are built into the web-based data collection instrument to detect major reporting errors. The system automatically generates percentages for many data elements, and totals for each survey page. Based on these calculations, edit checks compare current responses to previously reported data. The percent variance necessary to trigger an edit check varies depending on the data element being compared, but typically are considered out of the expected range if the variance is greater than 25 percent. Edit checks can be run by the keyholder at any time during the collection, and all edit failures are required to be resolved before the keyholder can lock the data. As edit checks are executed, survey respondents are allowed to correct any errors detected by the system. If data are entered correctly but fail the edit checks, the survey respondents are asked either to confirm that the data are correct as entered or to key in a text message explaining why the data appear to be out of the expected data range. Additionally, some edit failures are “fatal”; in these cases, the data must be corrected by the keyholder rather than confirmed or explained, or an edit override must be performed. Survey respondents are also provided with a context box for each survey component and are encouraged to use this area to explain any special circumstances that might not be evident in their reported data.

Final quality control procedures are performed when all institutions have responded or data for them have been imputed.

Before the conversion to a web-based reporting system, all data, whether received on paper forms, diskettes, electronically through the Internet, or through PETS, went through the same editing process to verify internal and inter-year consistency. Addition checks were performed by adding down or across columns and comparing generated totals with reported totals. If the reported total differed from the generated total but was within a designated range, the reported total was replaced by the generated total and the cell was flagged with the proper imputation code. Otherwise, institutions were contacted to resolve the discrepancies.

Estimation Methods

Imputation is done to compensate for nonresponding institutions—both those with total nonresponse and those with partial nonresponse to specific data items.

Prior to 1993, all sectors were surveyed and a sample of private less-than-2-year institutions was conducted to obtain national estimates for fall enrollment, completions, finance, and fall staff; these data were weighted and subject to sampling error. Starting in 1993, the IPEDS eliminated the sample of private less-than-2-year institutions and surveyed the entire universe of postsecondary institutions; therefore, no weighting is conducted.

Imputation. Impuation is performed after all editing has been completed. Several methods of imputation are used, depending on the availability of prior-year data, including a "carry forward" method, group means, and "nearest neighbor." All the IPEDS components use the same imputation flags. Institutions whose data are entirely imputed may be identified in the file by their response status and imputation type codes. For responding institutions whose data are partially imputed, the affected items may be identified by the associated item imputation flags.

In the past, the IPEDS used cold-deck (updated by ratio methods to reflect the change) and hot-deck imputation procedures to adjust for partial or total nonresponse to a specific survey instrument.

Recent Changes

Key changes to the IPEDS program since 1995 are summarized below:

- The primary focus of the IPEDS data collections is to collect data from Title IV institutions. These institutions have Program Participation Agreements (PPAs) with the Office of Postsecondary Education (OPE) within the U.S. Department of Education and thus are eligible to participate in Title IV student financial aid programs. The IPEDS program no longer differentiates between accredited college-level institutions and postsecondary institutions with occupational or vocational accreditation. Beginning with the 1996 data files, institutions have been classified by whether or not they are eligible to participate in Title IV financial aid programs and whether or not they grant degrees, not by highest level of offering.

- Between 1993 and 1996, NCES began to examine the universe of accredited institutions in order to form a crosswalk between the IPEDS data files and those maintained by OPE for student financial aid purposes. During this period, OPE discontinued its policy of differentiating institutions by level of accreditation—that is, those accredited at the college level (formerly the HEGIS universe) versus those with occupational/vocational accreditation. Since the IPEDS could no longer identify institutions with college-level accreditation, a new approach was developed to categorize institutions for mailout and analysis purposes. Beginning with the 1997 mailout, the IPEDS universe was subdivided according to (1) accreditation status, (2) level of institution, and (3) degree-granting status.

- Prior to the development of the web-based data collection system, the IPEDS survey forms were mailed to institutions based upon the information provided in the prior year's Institutional Characteristics Survey—control and highest level of offering (which determined an institution's sector) combined with accreditation status. Institutions that were not accredited, and thus not eligible for federal student financial aid, were asked to complete only the Institutional Characteristics survey form. All accredited institutions that either (1) grant an associate's or higher degree or (2) offer a certificate program above the baccalaureate level received a full packet of components—Institutional Characteristics, Completions, Fall Enrollment, Fall Enrollment in Occupationally Specific Programs, Fall Staff, Finance, Graduation Rates, Salaries of Full-Time Instructional Faculty, and Academic Libraries. All other accredited institutions (i.e., those granting only certificates at the subbaccalaureate level) were required to complete Institutional Characteristics,

Graduation Rates, and a Consolidated Form. In 2000, the IPEDS was redesigned, and postsecondary institutions that had Title IV Program Participation Agreements with OPE became the primary focus for the full set of data collected by the IPEDS. Thus, the current web-based system considers Title IV status rather than accreditation.

- In 1997, the Graduation Rate component was added to the IPEDS program to help institutions satisfy the requirements of the Student Right-to-Know Act of 1990.

- In 1999, NCES collected selected data items in a pilot test of a web-based survey. These items—tuition and fees for entering students, room and board, books and supplies, and information on students receiving financial aid—have been incorporated in the redesigned IPEDS data collection, implemented in 2000–01.

- In 2000–01, NCES converted the IPEDS to a web-based data collection system. The content of the survey "forms" was revised and reduced in scope, and the procedures for collecting data vary considerably from those used in prior years. In the first year, two collection cycles were implemented: the fall 2000 cycle collected Institutional Characteristics and Completions data, and the spring 2001 cycle collected Enrollment, Student Financial Aid, Finance, and Graduation Rate data. In subsequent years, a winter cycle has been included to collect Human Resources data.

- In 2005–06, three survey components—Employees by Assigned Position, Salaries, and Fall Staff—were merged into the single Human Resources component to simplify reporting and ensure data consistency and accuracy. The IPEDS glossary and instructions were also restructured, based on the new design, to improve the consistency of reporting between surveys. A few survey items were also reorganized to be more logical in flow.

- Beginning with the 2009–10 IPEDS, a new component was added to the spring collection, called 200% Graduation Rates (GR200). This component collects data on the number of students in the cohort who completed their program within 200 percent of normal time. It is separate from the regular Graduation Rates (GRS) component.

- In 2009-10, numerous changes were made to reduce reporting burden for nondegree-granting institutions. These changes include elimination of items on IC; combining data collection on HR into a single section with consolidation of 4 primary occupational categories (instruction, research, public service, and combined); elimination of transfers-in and noncertificate-seeking student columns on EF; and vastly simplifying the finance reporting required of these institutions.

Future Plans
The IPEDS plans to continue with three separate data collections (fall, winter, and spring) in future years. Data items may be modified to better reflect current issues in postsecondary education as recommended by the IPEDS Technical Review Panel. The next data collection is scheduled for the fall of 2010.

5. DATA QUALITY AND COMPARABILITY

Data element definitions have been formulated and tested to be relevant to all providers of postsecondary education and consistent among components of the system. A set of data elements has been established to identify characteristics common to all providers of postsecondary education, and specific data elements have been established to define unique characteristics of different types of providers. Interrelationships among various components of the IPEDS have been formed to avoid duplicative reporting and to enhance the policy relevance and analytic potential of the data. Through the use of "clarifying" questions that ask what was or was not included in a reported count or total or the use of context notes that supplement the web collection, it is possible to address problems in making interstate and interinstitutional comparisons. Finally, specialized, but compatible, reporting formats have been developed for the different sectors of postsecondary education providers. This design feature accommodates the varied operating characteristics, program offerings, and reporting capabilities that differentiate postsecondary institutional sectors, while yielding comparable statistics for some common parameters of all sectors.

Sampling Error
Only the data collected prior to 1993 from a sample of private less-than-2-year institutions are subject to sampling error. With this one exception, the HEGIS and the IPEDS programs include the universe of applicable postsecondary institutions.

Nonsampling Error

The IPEDS data are subject to such nonsampling errors as errors of design, reporting, processing, nonresponse, and imputation. To the extent possible, these errors are kept to a minimum by methods built into the survey procedures.

The sources of nonsampling error in the IPEDS data vary with the survey instrument. In the Fall Enrollment component, the major sources of nonsampling error are classification problems, the unavailability of needed data, misinterpretation of definitions, and operational errors. Possible sources of nonsampling error in the Finance component include nonresponse, imputation, and misclassification. The primary sources of nonsampling error in the Completions component are differences between the NCES program taxonomy and taxonomies used by colleges, classification of double majors and double degrees, operational problems, and survey timing. A major source of nonsampling error in the Graduation Rates components is the correct identification of cohort students (full-time, first-time, degree/certificate-seeking undergraduates); for Human Resources, difficulties in classifying employees by primary occupation; for 12-month Enrollment, definitional difficulties with calculating instructional activity. For Student Financial Aid, institutions often must merge enrollment and financial aid databases, and face difficulties in placing students in the various groups for which data are collected.

Coverage error. Coverage error in the IPEDS is believed to be minimal. For institutions that are eligible for Title IV federal financial aid programs, coverage is almost 100 percent. Schools targeted as “possible adds” are identified from many sources, including a review of the PEPS file from OPE, a universe review done by state coordinators, and the institutions themselves.

Nonresponse error. Since 1993, all institutions entering into PPAs with the U.S. Department of Education are required by law to complete the IPEDS package of components. Therefore, overall unit and item response rates are quite high for all components for these institutions. Data collection procedures, including extensive email and telephone follow-up, also contribute to the high response rates. Imputation is performed to adjust for both partial and total nonresponse to a survey. Because response rates are so high, error due to imputation is considered small.

Unit nonresponse. Because Title IV institutions are the primary focus of the IPEDS and they are required to respond, overall response rates for Title IV institutions and administrative units are high. For example, the overall response rate in winter 2007-08 was 99.9 percent for the HR component. The response rates were also 99.9 percent for the individual required HR sections: Employees by Assigned Position, Fall Staff, and Salaries. Since the implementation of the web collection, Title IV institutional response rates for the various IPEDS surveys have ranged from about 89 percent to over 99 percent. (See chapter 11 for response rates for the Academic Libraries Survey.)

By sector, the response rates are highest for public 4-year or higher institutions and lowest for private for-profit institutions, especially less-than-2-year institutions. The 1994 Academic Libraries and the FY 95 Finance public-use data files are limited to IHEs because the response rate for postsecondary institutions not accredited at the collegiate level was quite low (74 percent in the Finance Survey and less than 50 percent in the Academic Libraries Survey).

Item nonresponse. Most participating institutions provide complete responses for all items. Telephone and email follow-up are used to obtain critical missing items.

Measurement error. NCES strives to minimize measurement error in the IPEDS data by using various quality control and editing procedures. New questionnaire forms or items are field tested and/or reviewed by experts prior to use. To minimize reporting errors in the Finance component, NCES uses national standards for reporting finance statistics. Wherever possible, definitions and formats in the Finance component are consistent with those in the following publications: *College and University Business Administration; Administrative Services, Financial Accounting and Reporting Manual for Higher Education; Audits of Colleges and Universities; and HEGIS Financial Reporting Guide*.

The classification of students appears to be the main source of error in the Enrollment component. Institutions have had problems in correctly classifying first-time freshmen, other first-time students, and unclassified students for both full-time and part-time categories. These problems occur most often at 2-year institutions (both public and private) and private 4-year institutions. In the 1977–78 HEGIS validation studies, misclassification led to an estimated overcount of 11,000 full-time students and an undercount of 19,000 part-time students. Although the ratio of error to the grand total was quite small (less than 1 percent), the percentage of errors was as high as 5 percent at student detail levels and even higher at certain aggregation levels. (See also “Data Comparability” below.)

Data Comparability

The definitions and instructions for compiling the IPEDS data have been designed to minimize comparability problems. However, survey changes necessarily occur over the years, resulting in some issues of comparability. Also, postsecondary education institutions vary widely, and hence, comparisons of data provided by individual institutions may be misleading. Specific issues related to the comparability of the IPEDS data are described below.

Imputation. Imputed data are on file for institutions with partial or total nonresponse. *Caution should be exercised when comparing institutions for which data have been imputed, since these data are intended for computing national totals and not intended to be an accurate portrayal of an institution's data. Users should also be cautious when making year-to-year enrollment comparisons by state.* In some cases, state enrollment counts vary between years as a result of imputation rather than actual changes in the reported enrollment data. To avoid misinterpretation, users should always check the response status codes of individual institutions to determine if a large proportion of data was imputed.

Classification of institutions. Beginning in 1996, the subset of the IPEDS institutions eligible to participate in Title IV federal financial student aid has been validated by matching the IPEDS universe with the PEPS file maintained by OPE. Previously, institutions were self-identified as aid-eligible from the list of IHEs and responses to the Institutional Characteristics component.

Fields of study. In analyzing Completions data by field of study, users must remember that the data are reported at the institution level, and represent programs, not schools, colleges, or divisions within institutions. For example, some institutions might have a few computer and information science programs organized and taught within a business school. However, for the IPEDS reporting purposes, the degrees are classified and counted within the computer and information science discipline.

Reporting periods. The data collected through the IPEDS components for any one year represent two distinct time periods. The Institutional Characteristics, Enrollment and Human Resources data represent an institution at one point in time, the fall of the school year. 12-month Enrollment, Student Financial Aid, Finance, and Completions data cover an entire 12-month period or fiscal year. Some indicators in NCES reports use fall data in conjunction with 12-month data, and readers should be cognizant of the differences in time periods represented.

Questionnaire changes. Over the years, the IPEDS survey forms have undergone revisions that may have an impact on data comparability. Users should consider the following:

- The 2008-09 data collection was the start of a 3-year phase-in to the reporting of the new, 1997 federal race and ethnicity categories. The new categories allow students and staff to identify themselves using two or more race categories. The transition to the new race and ethnicity categories will be complete for the 2011-12 data collection.

- The 2008-09 data collection was the start of a 2-year phase-in of the restructuring of the postbaccalaureate degree categories. As of the 2010-11 data collection, the first-professional degree and certificate categories were eliminated, and the doctor's degree category was expanded to three categories: research/scholarship, professional practice, and other. These changes reflect changes in graduate education over the years, and make it easier to distinguish research-focused doctor's degrees from professionally focused doctor's degrees.

- Accreditation information was collected on the IC until 2006-07, when the Office of Postsecondary Education opened its database and searchable web tool of accredited institutions, collecting data from the accreditation agencies (http://ope.ed.gov/accreditation/).

- From 1990 to 1994, racial/ethnic data (by gender and degree/award level) were collected at the 2-digit CIP level on the Completions component. In 1995, there was a major restructuring of the component to collect race/ethnicity at the 6-digit CIP level and to add additional questions to collect numbers of completers with double majors and numbers of degrees granted at branch campuses in foreign countries. The additional questions were dropped in 2000–01, but a matrix to collect completions data on multiple majors was instituted for optional use in 2001–02 and became mandatory in 2002–03.

- Revisions to the CIP were made in 1970, 1980, 1985, 1990, 2000, and 2010. For a complete

history, please see **History of the Classification of Instructional Programs** later in this chapter.

- Racial/ethnic data for Fall Enrollment have been collected annually since 1990 (biennially, in even-numbered years, before then). Additional items were included on students enrolled in branch campuses in foreign countries, students enrolled exclusively in remedial courses, and students enrolled exclusively at extension divisions; however, these items were discontinued in 2000. Prior to 1996, data were also collected in even-numbered years from 4-year institutions for the fields of Veterinary Medicine and Architecture and Related Programs.

- Prior to 2000-01, the GRS collected additional data on students' length of time to complete; the number of students still persisting; and the number of students receiving athletically related student aid and their length of time to complete. The sections of the component collecting data on students receiving athletically related student aid were discontinued with the 2007-08 data collection.

- In 2009-10, forms used to collect Graduation Rates (GRS) data for less than 4-year institutions were modified to include reporting of completers within 100 percent of normal time in addition to 150 percent of normal time. This change aligned forms for the less than 4-year institutions with the 4-year institutions' forms.

- For the 2009–10 data collection, additional changes to the SFA component were implemented due to the Higher Education Opportunity Act (HEOA) and for clarification, including the collection of average aid amounts for sub-groups of the full-time, first-time degree/certificate-seeking undergraduate population, to be used in the calculation of average institutional net price and average institutional net price by income category information for display on the College Navigator website (http://nces.ed.gov/collegenavigator/).

- In fall 1995, the salary class intervals were revised for the Fall Staff component. Salary class intervals were revised again in 2001.

- Salary outlays, total number of instructional staff, and tenure status were collected for full-time staff on less than 9-month contract schedules through 1999-2000; currently only academic rank and gender are collected for these other contract schedules. Faculty status was not collected between 2001-02 and 2004-05, and was reinstated for degree-granting institutions in 2005-06. The reporting of data by faculty status was optional for 2005–06, but was required beginning in 2006–07. Beginning with the 2004-05 data collection, only degree-granting institutions have been required to complete the SA section of the HR component.

- As of the 2004-05 collection, the IPEDS has limited the collection of data on employees in medical schools to institutions with an M.D. or D.O. program. In previous collections, all 4-year institutions were given the opportunity to report employees in medical schools. However, some institutions that did not have a medical school erroneously reported employees in this section of the Employees by Assigned Position section. This change may cause some discrepancies in comparisons of the IPEDS medical school data.

- Prior to 2001, the Fall Staff component requested the number of persons donating (contributing) services or contracted for by the institution.

- Over the years, the various versions of the Finance form have changed. Prior to 1997, the survey forms for public and private institutions were basically the same except that the public institution form contained three additional sections, with data from questions pertaining to state and local government financial entities used by the U.S. Bureau of the Census.

- The Finance form for private institutions was revised in 1997 to make it easier for respondents to report their financial data according to new standards issued by the Financial Accounting Standards Board (FASB). In an attempt to address the reporting issues of proprietary institutions, the for-profit form was revised in 1999 to reflect the financial statements of these institutions. Due to new accounting standards issued by the Governmental Accounting Standards Board (GASB), beginning optionally in 2002, with a 2-year phase-in period, public GASB reporting

institutions moved from fund-based reporting to whole-entity reporting that is more similar to the private FASB-reporting institutions.

- With the web-based data collection, the number of data items requested from institutions was greatly reduced in FY 2000.

- A 2-year phase-in period began with FY 2008 reporting, to implement additional changes to better align the finance reporting of public and private institutions. For FY 2010 reporting, all public and not-for-profit institutions used the new aligned form.

History of Classification of Instructional Programs. The purpose of the Classification of Instructional Programs (CIP) is to provide a taxonomic scheme that supports the accurate tracking, assessment, and reporting of fields of study and program completions activity. NCES has utilized a number of versions of CIP throughout the life of IPEDS, as well as its predecessor, the Higher Education General Information System (HEGIS).

In 1970 NCES published "A Taxonomy of Instructional Programs in Higher Education" which was to be used beginning with the HEGIS surveys of 1971-72. This taxonomy was divided into two main sections: one dealt with conventional academic subdivisions of knowledge and training; the other with technologies and occupational specialties related to curricula leading to associate's degrees and other awards below the baccalaureate. Both sections used 4-digit numerical codes to represent the fields.

In 1981 NCES published "A Classification of Instructional Programs." In addition to new programs that evolved or gained new significance since 1970, there were weaknesses in the way instructional programs were classified and disaggregated. The new CIP instituted the current 6-digit code, which allowed obtaining data by 2-digit or 4-digit groups of fields more easily than the older scheme. The new CIP also included program definitions or descriptions, which the 1970 version lacked, as well as other improvements.

In 1985 another revision to the CIP was released, although this was more of an update to the 1980 CIP than a radical change. There were 116 fields deleted, either due to duplication, or because programs no longer existed to the degree needed for national reporting. Forty fields were added based on write-in entries on surveys returned. In addition, there were a few revisions of codes or names of fields. This CIP was used during the final years of HEGIS and continued into IPEDS.

A more extensive revision of CIP was released in 1990, which included programs at the secondary and adult education levels. Within the postsecondary level, there were several major restructures. Fields previously included in Business and Management (06) and Business (Administrative Support) (07) were integrated into a new Business Management and Administrative Support (52). Similarly, fields previously in Allied Health (17) and Health Sciences (18) were integrated into Health Professions and Related Sciences (51). Again there were deletions and additions, although many were actually combining two former fields into one, or vice versa. The 1990 CIP was first used in IPEDS 1991-92.

A further revision resulted in publishing "Classification of Instructional Programs: 2000 Edition" in 2002. This CIP was adopted as the standard field of study taxonomy by Statistics Canada, based on the comprehensiveness and detail of the CIP and the potential for enhanced comparability with U.S. education data. Again, there were several major reorganizations. Fields previously reported in Agricultural Sciences (02) were divided between Agriculture, Agriculture Operations and Related Sciences (01) and Biological and Biomedical Sciences (26). Fields previously reported in Sales and Marketing Operations/Marketing and Distribution (08) were incorporated into Business, Management, Marketing, and Related Services (52). History became a separate 2-digit CIP (54) moved from Social Sciences and History (45). In addition, there were a large number of new fields added. The CIP-2000 was first used in IPEDS in 2002-03.

The most recent revision to the CIP was developed during 2008-2009 and will be entirely on-line (http://nces.ed.gov/ipeds/cipcode/), with tools for browsing, searching, and crosswalking. There were fewer major shifts in coding; no new 2-digit series were added, and no large scale movement of codes from one series to another occurred. A large number of new fields were added: 50 new 4-digit codes and 300 new 6-digit codes. Several series were reorganized (English Language and Literature/Letters (23), Psychology (42), Nursing (51.16), and Residency Programs (60)), and one series was deleted (Technology Education/Industrial Arts (21)). Examples of instructional programs were added to assist users of CIP in selecting the appropriate field. This new version will be used in IPEDS for the 2010-11 collection.

Comparisons with HEGIS. *Caution must be exercised in making cross-year comparisons of institutional data collected in the IPEDS with data collected in HEGIS.* The IPEDS surveys request separate reporting by all institutions and their branches as long as each entity offers at least one complete program of study. Under HEGIS, only separately accredited branches of an institution were surveyed as separate entities; branches that were *not* separately accredited were combined with the appropriate entity for the purposes of data collection and reporting. Therefore, an institution may have several entities in the IPEDS, where only one existed in HEGIS.

Comparison with the Survey of Earned Doctorates. Like the IPEDS Completions Survey, the Survey of Earned Doctorates (SED) (see chapter 17) also collects data on doctoral degrees, but the information is provided by doctorate recipients rather than by institutions. The number of doctorates reported in the Completions component is slightly higher than in SED. This difference is largely attributable to the inclusion of nonresearch doctorates (primarily in theology and education) in the Completions component. The discrepancies in counts have been generally consistent since 1960, with ratios of the IPEDS-to-SED counts ranging from 1.01 to 1.06. Differences in the number of doctorates within a given field may be greater than the overall difference, because a respondent to SED may classify his or her specialty differently than how the institution reports the field in the Completions survey.

6. CONTACT INFORMATION

For content information on the IPEDS, contact:

Elise Miller
Phone: (202) 502-7318
E-mail: elise.miller@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Following each IPEDS collection cycle, three First Look publications are released. These publications present findings from the data collections, and include extensive survey methodology sections. They are available online. The latest three are listed below:

Knapp, L.G., Kelly-Reid, J.E., and Ginder, S.A. (2010). *Enrollment in Postsecondary Institutions, Fall 2008; Graduation Rates, 2002 and 2005 Cohorts; and Financial Statistics, Fiscal Year 2008* (NCES 2010-152REV). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Knapp, L.G., et al., Research Triangle Institute (2009). *Employees in Postsecondary Institutions, Fall 2008, and Salaries of Full-Time Instructional Staff, 2008-09* (NCES 2010-165). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Knapp, L.G. (2009). *Postsecondary Institutions and Price of Attendance in the United States: Fall 2008, Degrees and Other Awards Conferred: 2007-08, and 12-Month Enrollment: 2007-08* (NCES 2009-165). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Uses of Data

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2010). *Integrated Postsecondary Education Data System Glossary*. Retrieved from http://nces.ed.gov/ipeds/glossary/.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2010). *Classification of Instructional Programs*. Retrieved from http://nces.ed.gov/ipeds/cipcode/.

Data Quality and Comparability

Clery, S., Arntz, M., and Miller, A. (2008). Integrated Postsecondary Education Data System Human Resources Data Quality Study (NCES 2008-150). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Jackson, K.W., Jang, D., Sukasih, A., and Peeckson, S. (2005). *Integrated Postsecondary Education Data System Data Quality Study* (NCES 2005-175). National Center for Education Statistics, Institute of

Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 13: National Study of Postsecondary Faculty (NSOPF)

1. OVERVIEW

PERIODIC SURVEY OF A SAMPLE OF POSTSECONDARY INSTITUTIONS AND THEIR FACULTY

NSOPF includes:

- Institution Questionnaire
- Faculty Questionnaire
- Department Chairperson Questionnaire (1987–88 only)

The National Study of Postsecondary Faculty (NSOPF) is conducted to provide information on postsecondary faculty and instructional staff: their academic and professional background, sociodemographic characteristics, and employment characteristics, such as institutional responsibilities and workload, job satisfaction, and compensation. Thus far, there have been four NSOPF administrations—in the 1987–88 academic year (NSOPF:88), the 1992–93 academic year (NSOPF:93), the 1998–99 academic year (NSOPF:99), and the 2003–04 academic year (NSOPF:04). The first cycle was conducted with a sample of institutions, faculty, and department chairpersons. The second, third, and fourth cycles were limited to surveys of institutions and faculty, but with a substantially expanded sample of public and private, not-for-profit institutions and faculty. Furthermore, unlike any previous cycle of NSOPF, the fourth cycle was conducted in tandem with another study, the 2004 National Postsecondary Student Aid Study (NPSAS:04) (see chapter 14), as a component of a larger study, the 2004 National Study of Faculty and Students (NSoFaS:04).

Purpose

To provide a national profile of postsecondary faculty and instructional staff: their professional backgrounds, responsibilities, workloads, salaries, benefits, and attitudes.

Components

NSOPF consists of two questionnaires: one for institutions and one for faculty and instructional staff. Institutions receive both an *Institution Questionnaire* and a request to provide a faculty list. The *Faculty Questionnaire* is sent to faculty and instructional staff sampled from the lists provided by the institutions. The 1987–88 NSOPF also included a *Department Chairperson Questionnaire*.

Institution Questionnaire. The Institution Questionnaire obtains information on the number of full- and part-time instructional and noninstructional faculty (as well as instructional personnel without faculty status); the tenure status of faculty members (based on definitions provided by the institution); institution tenure policies (and changes in policies on granting tenure to faculty members); the impact of tenure policies on the influx of new faculty and on career development; the growth and promotion potential for existing nontenured junior faculty; the benefits and retirement plans available to faculty; and the turnover rate of faculty at the institution. The questionnaire is completed by an Institution Coordinator (IC) designated by the Chief Administrator (CA) at each sampled institution.

Faculty Questionnaire. This questionnaire addresses the following issues as they relate to postsecondary faculty and instructional staff: background characteristics

and academic credentials; workloads and time allocation between classroom instruction and other activities such as research, course preparation, consulting, public service, doctoral or student advising, conferences, and curriculum development; compensation and the importance of other sources of income, such as consulting fees and royalties; the role of faculty in institutional policymaking and planning (and the differences, if any, between the role of part- and full-time faculty); faculty attitudes toward their jobs, their institutions, higher education, and student achievement in general; changes in teaching methods and the impact of new technologies on teaching techniques; career and retirement plans; differences between individuals who have instructional responsibilities and those who do not (e.g., those engaged only in research); and differences between those with teaching responsibilities but no faculty status and those with teaching responsibilities and faculty status. Eligible respondents for this questionnaire are faculty and instructional staff sampled from lists provided by institutions involved in the study. These lists are compiled by the IC at each sampled institution.

Department Chairperson Questionnaire. Administered only in the 1987–88 academic year, this questionnaire collected information from over 3,000 department chairpersons on the faculty composition in departments, tenure status of faculty, faculty hires and departures, hiring practices, activities used to assess faculty performance, and professional and developmental activities.

Periodicity

The NSOPF was conducted in the 1987–88, 1992–93, 1998–99, and 2003–04 academic years. No specific administration date has been set for the next round of NSOPF.

2. USES OF DATA

NSOPF provides valuable data on postsecondary faculty that can be applied to policy and research issues of importance to federal policymakers, education researchers, and postsecondary institutions across the United States. For example, NSOPF data can be used to analyze whether the size of the postsecondary labor force is decreasing or increasing. NSOPF data can also be used to analyze faculty job satisfaction and how it correlates with an area of specialization as well as how background and specialization skills relate to present assignments. Comparisons can be made on academic rank and outside employment. Benefits and compensation can be studied across institutions, and faculty can be aggregated by sociodemographic characteristics. Because NSOPF is conducted periodically, it also supports comparisons of data longitudinally.

The Institution Questionnaire includes items about

- the number of full- and part-time faculty (regardless of whether they had instructional responsibilities) and instructional personnel without faculty status;
- the distribution of faculty and instructional staff by employment (i.e., full-time, part-time) and tenure status (based on the definitions provided by the institution);
- institutional tenure policies and changes in policies on granting tenure to faculty members;
- the impact of tenure policies on the number of new faculty and on career development;
- the growth and promotion potential for existing nontenured junior faculty;
- the procedures used to assess the teaching performance of faculty and instructional staff;
- the benefits and retirement plans available to faculty; and
- the turnover rates of faculty at the institution.

The Faculty Questionnaire addresses such issues as respondents' employment, academic, and professional background; institutional responsibilities and workload; job satisfaction; compensation; sociodemographic characteristics; and opinions. The questionnaire is designed to emphasize behavioral rather than attitudinal questions in order to collect data on who the faculty are; what they do; and whether, how, and why the composition of the nation's faculty is changing.

The Faculty Questionnaire includes items about

- background characteristics and academic credentials;
- workloads and time allocation between classroom instruction and other activities (such as research, course preparation, consulting, work at other institutions, public service,

doctoral or student advising, conferences, and curriculum development);

- compensation and the importance of other sources of income, such as consulting fees and royalties;
- the number of years spent in academia, and the number of years with instructional responsibilities;
- the role of faculty in institutional policymaking and planning (and the differences, if any, between the role of full- and part-time faculty);
- faculty attitudes toward their jobs, their institutions, higher education, and student achievement in general;
- changes in teaching methods, and the impact of new technologies on instructional techniques;
- career and retirement plans;
- differences between those who have instructional responsibilities and those who do not, such as those engaged only in research; and
- differences between those with teaching responsibilities but no faculty status and those with teaching responsibilities and faculty status.

3. KEY CONCEPTS

Some key concepts related to NSOPF are described below.

Faculty/Instructional Staff (NSOPF:04). Eligible individuals for NSOPF:04 included any faculty and instructional staff who

- were permanent, temporary, adjunct, visiting, acting, or postdoctoral appointees;
- were employed full- or part-time by the institution;
- taught credit or noncredit classes;
- were tenured, nontenured but on a tenure track, or nontenured and not on a tenure track;
- provided individual instruction, served on thesis or dissertation committees, or advised or otherwise interacted with first-professional, graduate, or undergraduate students;
- were in professional schools (e.g., medical, law, or dentistry); or
- were on paid sabbatical leave.

NSOPF:04 excluded staff who

- were graduate or undergraduate teaching or research assistants;
- had instructional duties outside of the United States, unless on sabbatical leave;
- were on leave without pay;
- were not paid by the institution (e.g., those in the military or part of a religious order);
- were supplied by independent contractors; or
- otherwise volunteered their services.

Faculty/Instructional Staff (NSOPF:99).
Faculty—All employees classified by the institution as faculty who were on the institution's payroll as of November 1, 1998. Included as faculty were

- any individuals who would be reported as "Faculty (Instruction/Research/Public Service)" in the U.S. Department of Education's 1997–98 Integrated Postsecondary Education Data System (IPEDS) Fall Staff Survey[1] (see chapter 12);
- any individuals with faculty status who would be reported as "Executive, Administrative, and Managerial" in the 1997–98 IPEDS Fall Staff Survey, whether or not they engaged in any instructional activities; and
- any individuals with faculty status who would be reported as "Other Professionals (Support/Service)" in the 1997–98 IPEDS Fall Staff Survey, whether or not they engaged in any instructional activities.

[1] When constructing the NSOPF:99 institution frame, faculty data from 1995–96 IPEDS were used if 1997–98 data were missing.

Individuals who would be reported as “Instruction/Research Assistants” in the 1997–98 IPEDS Fall Staff Survey were excluded.

Instructional Staff—All employees with instructional responsibilities—those teaching one or more courses, or advising or supervising students’ academic activities (e.g., by serving on undergraduate or graduate thesis or dissertation committees or supervising an independent study or one-on-one instructions)—who may or may not have had faculty status. Included as instructional staff were

- any individuals with instructional responsibilities during the 1998 fall term who would be reported as “Executive, Administrative, and Managerial” in the 1997–98 IPEDS Fall Staff Survey (e.g., a finance officer teaching a class in the business school); and

- any individuals with instructional responsibilities during the 1998 fall term who would be reported as “Other Professionals (Support/Service)” in the 1997–98 IPEDS Fall Staff Survey.

Individuals who would be reported as “Instruction/Research Assistants” in the 1997–98 IPEDS Fall Staff Survey were excluded.

Faculty/Instructional Staff (NSOPF:93). All institutional staff (faculty and nonfaculty) whose major regular assignment at the institution (more than 50 percent) was instruction. This corresponds to the definition used in IPEDS glossary (Broyles 1995), which defines faculty (instruction/research/public service) as “persons whose specific assignments customarily are made for the purpose of conducting instruction, research, or public service as a principal activity (or activities), and who hold academic-rank titles of professor, associate professor, assistant professor, instructor, lecturer, or the equivalent of any of these academic ranks. If their principal activity is instructional, this category includes deans, directors, or the equivalent, as well as associate deans, assistant deans, and executive officers of academic departments…”

A dedicated instructional assignment was not required for an individual to be designated as faculty/instructional staff in NSOPF:93. Included in the definition were: administrators whose major responsibility was instruction; individuals with major instructional assignments who had temporary, adjunct, acting, or visiting status; individuals whose major regular assignment was instruction but who had been granted release time for other institutional activities; and individuals whose major regular assignment was instruction but who were on sabbatical leave from the institution. Excluded from this definition were graduate or undergraduate teaching assistants, postdoctoral appointees, temporary replacements for personnel on sabbatical leave, instructional personnel on leave without pay or teaching outside the United States, military personnel who taught only Reserve Officers Training Corps (ROTC) courses, and instructional personnel supplied by independent contractors.

Noninstructional Faculty (NSOPF:93). All institutional staff who had faculty status but were not counted as instructional faculty since their specific assignment was *not instruction* but rather conducting research, performing public service, or carrying out administrative functions.

Instructional Faculty (NSOPF:88). Those members of the institution’s instruction/research staff who were employed full- or part-time (as defined by the institution) and whose assignment included instruction. Included were administrators, such as department chairs or deans, who held full- or part-time faculty rank and whose assignment included instruction; regular full- and part-time instructional faculty; individuals who contributed their instructional services, such as members of religious orders; and instructional faculty on sabbatical leave. Excluded from this definition were teaching assistants; replacements for faculty on sabbatical leave; faculty on leave without pay; and others with adjunct, acting, or visiting appointments.

4. SURVEY DESIGN

Target Population

Since NSOPF:99, the target population has consisted of all public and private, not-for-profit Title IV-participating, 2- and 4-year degree-granting institutions in the 50 states and the District of Columbia that offer programs designed for high school graduates and are open to persons other than employees of the institution and faculty and instructional staff in these institutions. The NSOPF:93 and NSOPF:88 institution-level population included postsecondary institutions with accreditation at the college level recognized by the U.S. Department of Education. The NSOPF:88 faculty-level population included only instructional faculty, but it also targeted department chairpersons.

Sample Design

NSOPF:04 used a two-stage sample design, with a sample of 1,080 institutions selected for participation in the first stage, of which 1,070 were eligible and 890 provided a faculty list suitable for sampling. In the second stage, a total of 35,630 faculty were sampled from participating institutions. Of these, 34,330 were eligible.

The institution frame was constructed from the Winter 2001–02 IPEDS data file. Institutions were partitioned into institutional strata based on institutional control, highest level of offering, and Carnegie classification. The sample of institutions was selected with probability proportional to size (PPS) based on the number of faculty and students at each institution.

In the faculty-level stage of sampling, faculty were grouped into strata based on race/ethnicity, gender, and employment status. Furthermore, the faculty sample was implicitly stratified by academic field. Stratifying the faculty in this way allowed for the oversampling of relatively small subpopulations (such as members of Black, Hispanic, and other ethnic/racial groups) in order to increase the precision of the estimates for these groups. The selection procedure allowed the sample sizes to vary across institutions, but minimized the variation in the weights within the staff-level strata: the sampling fractions for each sample institution were made proportional to the institution weight.

The sample for NSOPF:99 was selected in three stages. Both the first-stage sample of institutions and the second-stage sample of faculty were stratified, systematic samples. In the initial stage, 960 postsecondary institutions were selected from the 1997–98 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) data files and the 1997 and 1995 IPEDS Fall Staff files. Each sampled institution was asked to provide a list of all of the full- and part-time faculty that the institution employed during the 1998 fall term, and 819 institutions provided such a list. In the second stage of sampling, some 28,580 faculty were selected from the lists provided by the institutions. Over 1,500 of these sample members were determined to be ineligible for NSOPF:99, as they were not employed by the sampled institution during the 1998 fall term, resulting in a sample of 27,040 faculty. A third stage of sampling occurred in the final phases of data collection. In order to increase the response rate and complete data collection in a timely way, a subsample of the faculty who had not responded was selected for intensive follow-up efforts. Others who had not responded were eliminated from the sample, resulting in a final sample of 19,210 eligible faculty.

NSOPF:93 was conducted with a sample of 970 postsecondary institutions (public and private, not-for-profit 2- and 4-year institutions whose accreditation at the college level was recognized by the U.S. Department of Education) in the first stage and 31,350 faculty sampled from institution faculty lists in the second stage. Institutions were selected from IPEDS and then classified into 15 strata by school type, based on their Carnegie Classifications. The strata were (1) private, other Ph.D. institution (not defined in any other stratum); (2) public, comprehensive; (3) private, comprehensive; (4) public, liberal arts; (5) private, liberal arts; (6) public, medical; (7) private, medical; (8) private, religious; (9) public, 2-year; (10) private, 2-year; (11) public, other type (not defined in any other stratum); (12) private, other type (not defined in any other stratum); (13) public, unknown type; (14) private, unknown type; and (15) public, research; private, research; and public, other Ph.D. institution (not defined in any other stratum). Within each stratum, the institutions were further sorted by school size. Of the 960 eligible institutions, 820 (85 percent) provided lists of faculty. The selection of faculty within each institution was random except for the oversampling of the following groups: Blacks (both non-Hispanics and Hispanics); Asians/Pacific Islanders; faculty in disciplines specified by the National Endowment for the Humanities; and full-time female faculty.

NSOPF:88 was conducted with a sample of 480 institutions (including 2-year, 4-year, doctoral-granting, and other colleges and universities), some 11,010 faculty, and more than 3,000 department chairpersons. Institutions were sampled from the 1987 IPEDS universe and were stratified by modified Carnegie Classifications and size (faculty counts). These strata were (1) public, research; (2) private, research; (3) public, other Ph.D. institution (not defined in any other stratum); (4) private, other Ph.D. institution (not defined in any other stratum); (5) public, comprehensive; (6) private, comprehensive; (7) liberal arts; (8) public, 2-year; (9) private, 2-year; (10) religious; (11) medical; and (12) “other” schools (not defined in any other stratum). Within each stratum, institutions were randomly selected. Of the 480 institutions selected, 450 (94 percent) agreed to participate and provided lists of their faculty and department chairpersons. Within 4-year institutions, faculty and department chairpersons were stratified by program area and randomly sampled within each stratum; within 2-year institutions, simple random samples of faculty and department chairpersons were selected; and within specialized institutions (religious, medical, etc.), faculty samples were randomly selected (department chairpersons were not sampled). At all institutions, faculty were also stratified on the basis of

employment status—full-time and part-time. Note that teaching assistants and teaching fellows were excluded in NSOPF:88.

Data Collection and Processing

NSOPF:04 allowed ICs to upload lists of faculty and instructional staff and to complete the Institution Questionnaire online. Institutions were also given the option of responding by telephone, though a web response was preferred. Faculty and instructional staff were allowed to participate via a self-administered web-based questionnaire or an interviewer-administered telephone interview (CATI). Follow-up with ICs and with faculty was conducted by telephone, mail, and e-mail.

NSOPF:99 allowed sample members to complete a self-administered paper questionnaire and mail it back or to complete the questionnaire online. Follow-up activities included e-mails, telephone prompting, and, for nonresponding faculty, CATI. As part of the study, an experiment was conducted to determine if small financial incentives could increase use of the web-based version of the questionnaire. Previously, NSOPF was a mailout/mailback survey with telephone follow-up.

NSOPF:88 was conducted by SRI International; NSOPF:93 by the National Opinion Research Center (NORC) at the University of Chicago; NSOPF:99 by The Gallup Organization; and NSOPF:04 by RTI International.

Reference Dates. Most of the information collected in NSOPF pertains to the fall term of the academic year surveyed. For NSOPF:04, the fall term was defined as the academic term containing November 1, 2003. The Institution Questionnaire also asked about the number of full-time faculty/instructional staff considered for tenure in the 2003–04 academic year. The NSOPF:04 Faculty Questionnaire asked faculty and instructional staff about the year they began their first faulty or instructional staff position at a postsecondary institution; the number of presentations and publications during their entire career and, separately, the number during the last 2 years; and their gross compensation and household income in calendar year 2003. Similarly, NSOPF:99, NSOPF:93, and NSOPF:88 requested most information for the 1998, 1992, and 1987 fall term, respectively, but included some questions requiring retrospective or prospective responses.

Data Collection. The NSOPF:04 data collection offered both a CATI and a web-based version of the Institution and Faculty questionnaires, with mail, telephone, and e-mail follow-up. Some 1,070 institutions in the eligible institution sample for the 2004 National Study of Faculty and Students (NSoFaS:04) were sampled and recruited to participate in both components of NSoFaS:04 (NSOPF:04 and NPSAS:04). The fielding of NSOPF:04 and NPSAS:04 together as NSoFaS:04 was one of three changes made in the institution contacting procedures for this cycle of NSOPF. The second change was to administer the Institution Questionnaire as a web or CATI instrument, with no hard-copy equivalent. The third change was to begin recruiting institutions and initiating coordinator contacts in March 2003—a full 8 months prior to the November reference date for the fall term and 5 to 6 months earlier than the September start dates of previous cycles. This change was prompted by the need to draw a faculty sample and subsequently contact sampled faculty for participation prior to the 2004 summer break.

The data collection procedure started in March 2003 with a cover letter and a set of pamphlets on NSoFaS, NSOPF, and NPSAS being sent to the institution's Chief Administrator (CA) as an introduction to the study. Study personnel then followed up with the CA by telephone, asking him or her to name an IC. An information packet was then sent to the IC. Each IC was then asked to complete a Coordinator Response Form to confirm that the institution could supply the faculty list within stated schedule constraints. ICs who indicated that a formal review process was needed before their institution would participate were forwarded additional project materials as appropriate.

A binder containing complete instructions for NSOPF:04, as well as a request for a faculty/instructional staff list, was sent to ICs in September 2003. ICs were asked to complete the Institution Questionnaire using the study's website. Data collection for the Institution Questionnaire ended in October 2004.

In NSOPF:04 full-scale study, the faculty data collection began with introductory materials being sent to sample members via first-class mail as well as e-mail. The letter included instructions for completing the self-administered questionnaire on the Internet or by calling a toll-free number to complete a telephone interview. After an initial 4-week period, telephone interviewers began calling sample members. An early-response incentive, designed to encourage sample members to complete the self-administered questionnaire prior to outgoing CATI calls, was offered to sample members who completed the questionnaire within 4 weeks of the initial mailing. Incentives were also offered to selected sample members as necessary

(i.e., those who refused to complete the questionnaire and other nonrespondents).

The NSOPF:99 data collection offered both a paper and a web version of the Institution and Faculty questionnaires, with telephone (including CATI) and e-mail follow-up. The data collection procedure started with a prenotification letter to the institution's CA to introduce him or her to the study and secure the name of an appropriate individual to serve as the IC. The data collection packet was then mailed directly to the IC. The packet contained both the Institution Questionnaire and the faculty list collection packet. The IC was asked to complete and return all materials at the same time. The mailing was timed to immediately precede the November 1, 1998, reference date for the fall term.

The field period for the NSOPF:99 faculty data collection extended from February 1999 through March 2000. Questionnaires were mailed to faculty in waves, as lists of faculty and instructional staff were received, processed, and sampled. Questionnaires were accompanied by a letter that provided the web address and a unique access code to be used to access the web questionnaire. The first wave of questionnaires was mailed on February 4, 1999; the seventh and final wave was mailed on December 1, 1999. Faculty sample members in each wave received a coordinated series of mail, e-mail, and telephone follow-ups. Mail follow-up for nonrespondents included a postcard and up to four questionnaire re-mailings; these were mailed to the home address of the faculty member if provided by the institution. E-mail prompts were sent to all faculty for whom an e-mail address was provided; faculty received as many as six e-mail prompts. Telephone follow-up consisted of initial prompts to complete the mail or web questionnaire. A CATI was scheduled for nonrespondents to the mail, e-mail, and telephone prompts.

The following efforts were made for the NSOPF:93 institution data collection: initial questionnaire mailing, postcard prompting, second questionnaire mailing, second postcard prompting, telephone prompting, third questionnaire mailing, and telephone interviewing. Similarly, the NSOPF:93 faculty data collection used an initial questionnaire mailing, postcard prompting, second questionnaire mailing, third questionnaire mailing, telephone prompting, and CATI. In both collections, institutions and faculty who missed critical items and/or had inconsistent or out-of-range responses were identified for data retrieval. Extra telephone calls were made to retrieve these data.

Data collection procedures for NSOPF:88 involved three mailouts for both the Institution Questionnaire and the Department Chairperson Questionnaire, and two mailouts and one CATI interview for the Faculty Questionnaire.

Data Processing. The NSoFaS:04 website was used for both NSOPF:04 and NPSAS:04. For institutions, it was a central repository for all study documents and instructions. It allowed for the uploading of electronic lists of faculty and instructional staff. In addition, it housed the Institution Questionnaire for the IC to complete online.

For NSOPF:04, institutions were asked to provide a single, unduplicated (i.e., with duplicate entries removed) electronic list of faculty in any commonly used and easily processed format (e.g., ASCII fixed field, comma delimited, spreadsheet format). However, as in previous cycles, paper lists were accepted, as were multiple files (e.g., separate files of full- and part-time faculty) and lists in electronic formats that did not lend themselves to electronic processing (such as word processing formats). For the first time, institutions were given the option of transmitting their electronic faculty lists via a secure upload to the NSoFaS:04 website and were encouraged to do so. (In previous cycles, direct upload was available only by file-transfer protocols, an option that few institutions utilized.) Institutions were also given the option of sending a CD-ROM or diskette containing the list data or sending the list via e-mail (as an encrypted file, if necessary).

Follow-up with ICs was conducted by telephone, mail, and e-mail. As faculty lists were received, they were reviewed for completeness, readability, and accuracy. Additional follow-up to clarify the information provided or retrieve missing information was conducted by the institution contactors as necessary. For institutions lacking the resources to provide a complete list of full- and part-time faculty and instructional staff, list information was, if possible, abstracted from course catalogs, faculty directories, and other publicly available sources. Faculty lists abstracted in this fashion were reviewed for completeness against IPEDS before being approved for sampling.

Institution Questionnaire follow-up was conducted simultaneously with follow-up for lists of faculty. If an institution was unable to complete the questionnaire online, efforts were made to collect the information by telephone. To expedite data collection, missing questionnaire data was, in some instances, abstracted directly from benefits and policy documentation supplied by the institution or from information publicly available on the institution's website.

For the faculty data collection, NSOPF:04 also utilized a mixed-mode data collection methodology that allowed sample members to participate via a web-based self-administered questionnaire or via CATI. The NSOPF:04 faculty instrument was designed to minimize potential mode effects by using a single instrument for both self-administration and CATI interviews. Four weeks after the release of the web-based questionnaire, nonrespondents were followed up to conduct a CATI interview.

Faculty lists and questionnaire data were evaluated by the project staff for quality, item nonresponse, item mode effects, break-offs, coding, quality control monitoring of interviewers, and interviewer feedback.

In NSOPF:99, each of the three modes of questionnaire administration required separate systems for data capture. All self-administered paper questionnaires were optically scanned. The system was programmed so that each character was read and assigned a confidence level. All characters with less than a 100 percent confidence level were automatically sent to an operator for manual verification. The contractor verified the work of each operator and the recognition engines on each batch of questionnaires to ensure that the quality assurance system was working properly. Also, 100 percent of written-out responses (as opposed to check marks) were manually verified.

Each web respondent was assigned a unique access code, and respondents without a valid access code were not permitted to enter the website. A respondent could return to the survey website at a later time to complete a survey that was left unfinished in an earlier session. When respondents entered the website using the access code, they were immediately taken to the same point in the survey item sequence that they had reached during their previous session. If respondents, re-using an access code, returned to the website at a later time after completing the survey in a previous session, they were not allowed access to the completed web survey data record. Responses to all web-administered questionnaires underwent data editing, imputation, and analysis.

All telephone interviews used CATI technology. The CATI program was altered from the paper questionnaire to ensure valid codes, perform skip patterns automatically, and make inter-item consistency checks where appropriate. The quality control program for CATI interviewing included project-specific training of interviewers, regular evaluation of interviewers by interviewing supervisors, and regular monitoring of interviewers.

NSOPF:93 used both computer-assisted data entry (CADE) and CATI. The CADE/CATI systems were designed to ensure that all entries conformed to valid ranges of codes; enforce skip patterns automatically; conduct inter-item consistency checks, where appropriate; and display the full question-and-answer texts for verbatim responses. As part of the statistical quality control program, 100 percent verification was conducted on a randomly selected subsample of 10 percent of all Institution and Faculty questionnaires entered in CADE. The error rate was less than 0.5 percent for all items keyed. Quality assurance for CATI faculty interviews consisted of random online monitoring by supervisors.

Editing and Coding. For the study in general, a large part of the data editing and coding was performed in the data collection instruments, including range edits; across-item consistency edits; and coding of fields of teaching, scholarly activities, and highest degree. During and following data collection, the data were reviewed to confirm that the data collected reflected the intended skip-pattern relationships. At the conclusion of the data collection, special codes were inserted in the database to reflect the different types of missing data.

The data cleaning and editing process in NSOPF:04 consisted of the following steps:

(1) *Review of one-way frequencies for every variable to confirm that there were no missing or blank values and to check for reasonableness of values.* This involved replacing blank or missing data with -9 for all variables in the instrument database and examining frequencies for reasonableness of data values.

(2) *Review of two-way cross-tabulations between each gate-nest combination of variables to check data consistency*. Gate variables are items that determine subsequent instrument routing. Nest variables are items that are asked or not asked, depending on the response to the gate question. Legitimate skips were identified using the interview programming code as specifications to define all gate-nest relationships and replace -9 (missing values that were blank because of legitimate skips) with -3 (legitimate skip code). Additional checks ensured that the legitimate skip code was not overwriting valid data and that no skip logic was missed. In addition, if a gate variable was missing (-9), the -9 was carried through the nested items.

(3) *Identify and code items that were not administered due to a partial or abbreviated interview*. This code replaced -9 values with -7 (item not administered) based on the section completion and abbreviated interview indicators.

(4) *Recode "don't know" responses to missing.* This code replaced -1 (don't know) values with -9 (missing) for later stochastic imputation. For selected items for which "don't know" seemed like a reasonable response, variables were created both with and without the "don't know" category.

(5) *Identify items requiring recoding.* During this stage, previously uncodable values (e.g., text strings) collected in the various coding systems were upcoded, if possible.

(6) *Identify items requiring range edits, logical imputations, and data corrections.* Descriptive statistics for all continuous variables were examined. Values determined to be out-of-range were either coded to the maximum (or minimum) reasonable value or set to missing for later imputation. Logical imputations were implemented to assign values to legitimately skipped items whose values could be implicitly determined from other information provided. Data corrections were performed where there were inconsistencies between responses given by the sample member.

Estimation Methods

Weighting was used in NSOPF to adjust for sampling and unit nonresponse at both the institution and faculty levels. Imputation was performed to compensate for item nonresponse.

Weighting. In NSOPF:04, three weights were computed: full-sample institution weights, full-sample faculty weights, and a contextual weight (to be used in "contextual" analyses that simultaneously include variables drawn from the Faculty and Institution questionnaires). The formulas representing the construction of each of these weights are provided in the *2004 National Study of Postsecondary Faculty (NSOPF:04) Methodology Report* (Huer et al. 2005). NSOPF:99 used weighting procedures similar to those used in NSOPF:04. For details on these procedures, see the *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (Abraham et al. 2002).

The weighting procedures used in NSOPF:93 and NSOPF:88 are described below.

NSOPF:93. Three weights were computed for the NSOPF:93 sample—first-stage institution weights, final institution weights, and final faculty weights. The first-stage institution weights accounted for the institutions that participated in the study by submitting a faculty list that allowed faculty members to be sampled. The two final weights—weights for the sample faculty and for institutions that returned the Institution Questionnaire—were adjusted for nonresponse. The final faculty weights were poststratified to the "best" estimates of the number of faculty. The "best" estimates were derived following reconciliation and verification through recontact with a subset of institutions that had discrepancies of 10 percent or more between the total number enumerated in their faculty list and Institution Questionnaire. For more information on the reconciliation effort, see "Measurement Error" (in section 5 below). For more information on the calculation of the "best" estimates of faculty, see the *1993 National Study of Postsecondary Faculty Methodology Report* (Selfa et al. 1997).

NSOPF:88. The NSOPF:88 sample was weighted to produce national estimates of institutions, faculty, and department chairpersons by using weights designed to adjust for differential probabilities of selection and nonresponse. The sample weights for institutions were calculated as the inverse of the probability of selection, based on the number of institutions in each size substratum. Sample weights were adjusted to account for nonresponse by multiplying the sample weights by the reciprocal of the response rate. Sample weights for faculty in NSOPF:88 summed to the total number of faculty in the IPEDS universe of institutions, as projected from the faculty lists provided by participating institutions, and accounted for two levels of nonresponse: one for nonparticipating institutions and one for nonresponding faculty. Sample weights for department chairpersons in NSOPF:88 summed to the estimated total number of department chairpersons in the IPEDS universe of institutions and accounted for nonresponse of nonparticipating institutions and nonresponding department chairpersons.

Imputation. Data imputation for the NSOPF:04 Faculty Questionnaire was performed in four steps:

(1) *Logical imputation.* The logical imputation was conducted during the data cleaning steps (as explained under "Editing and Coding" above).

(2) *Cold deck.* Missing responses were filled in with data from the sample frame or institution record data whenever the relevant data were available.

(3) *Sequential hot deck.* Nonmissing values were selected from "sequential nearest neighbors" within the imputation class. All questions that were categorical and had more than 16 categories were imputed with this method.

(4) *Consistency checks.* After all variables were imputed, consistency checks were applied to the entire faculty data file to ensure that the imputed values did not conflict with other questionnaire items, observed or imputed. This process involved reviewing all of the logical imputation and editing rules as well.

Data imputation for the institution questionnaire used three methods, within-class mean, within-class random frequency, and hot deck. The imputation method for each variable is specified in the labels for the imputation flags in the institution dataset. Logical imputation was also performed in the cleaning steps described previously in the "Editing and Coding" section.

Imputation for the NSOPF:99 Faculty Questionnaire was performed in four steps:

(1) *Logical imputation*. The logical imputation was conducted during the data cleaning steps (as explained under "Editing and Coding" above).

(2) *Cold deck*. Missing responses were filled in with data from the sample frame whenever the relevant data were available.

(3) *Sequential hot deck*. Nonmissing values were selected from "sequential nearest neighbors" within the imputation class. All questions that were categorical and had more than 16 categories were imputed with this method.

(4) *Regression type*. This procedure employed SAS PROC IMPUTE. All items that were still missing after the logical, cold-deck, and hot-deck imputation procedures were imputed with this method. Project staff selected the independent variables by first looking through the questionnaire for logically related items and then by conducting a correlation analysis of the questions against each other to find the top correlates for each item.

Data imputation for the NSOPF:99 Institution Questionnaire used three methods. Logical imputation was also performed in the cleaning steps described under "Editing and Coding."

(1) *Within-class mean*. The missing value was replaced with the mean of all nonmissing cases within the imputation class. Continuous variables with less than 5 percent missing data were imputed with this method.

(2) *Within-class random frequency*. The missing value was replaced by a random draw from the possible responses based on the observed frequency of nonmissing responses within the imputation class. All categorical questions were imputed with this method, since all categorical items had less than 5 percent missing data.

(3) *Hot deck*. As with the faculty imputation, this method selected nonmissing values from the "sequential nearest neighbor" within the imputation class. Any questions that were continuous variables and had more than 5 percent missing cases were imputed with this method.

For a small number of items, special procedures were used. See the *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (Abraham et al. 2002).

In NSOPF:93, two imputation methods were used for the Faculty Questionnaire—PROC IMPUTE and the "sequential nearest neighbor" hot-deck method. PROC IMPUTE alone was used for the NSOPF:93 Institution Questionnaire. All imputation was followed by a final series of cleaning passes that resulted in generally clean and logically consistent data. Some residual inconsistencies between different data elements remained in situations where it was impossible to resolve the ambiguity as reported by the respondent.

Although NSOPF:88 consisted of three questionnaires, imputations were only performed for faculty item nonresponse. The within-cell random imputation method was used to fill in most Faculty Questionnaire items that had missing data.

Recent Changes

NSOPF:04 was, in one respect, unlike any previous cycle of NSOPF, as it was conducted in tandem with another major study, NPSAS:04, under one overarching contract: NSoFaS:04. NCES recognized that, historically, there had been considerable overlap in the institutions selected for participation in

NSOPF:04 and NPSAS:04. By combining the two independent studies under one contract, NCES sought to minimize the response burden on institutions and to realize data collection efficiencies. Nevertheless, NSOPF:04 and NPSAS:04 retain their separate identities. The purpose of this chapter is to summarize the methodology of NSOPF:04; sampling and data collection procedures for NPSAS:04 are referred to only as they are combined with, or impact, the parallel procedures for NSOPF:04.

The combination of NSOPF:04 and NPSAS:04 into NSoFaS:04 had important implications for the NSOPF:04 institution sample design and institution contacting procedures. Institutions for the NSOPF:04 sample were selected as a subsample of the NPSAS:04 sample. This combination resulted in a somewhat larger sample of institutions for the full-scale study than in previous NSOPF cycles (1,070 eligible institutions in NSOPF:04 compared to 960 in NSOPF:99) and created a need to balance the design requirements of both studies in all institution-related study procedures.

Future Plans
A specific date has not yet been selected for the next administration of NSOPF.

5. DATA QUALITY AND COMPARABILITY

NSOPF:04 included procedures for both minimizing and measuring nonsampling errors. A field test was performed before NSOPF:04, and quality control activities continued during interviewer training, data collection, and data processing.

Sampling Error
Standard errors for all NSOPF data can be computed using a technique known as Taylor Series approximation. Individuals opting to calculate variances with the Taylor Series approximation method should use a “with replacement” type of variance formula. Specialized computer programs, such as SUDAAN, calculate variances with the Taylor Series approximation method. The Data Analysis System (DAS) from NCES available on CD-ROM calculates variances using the Taylor Series method, and the DAS available online calculates variances using the balanced repeated replicate method.

Replicate weights are provided in the NSOPF data files (64 sets of replicates in NSOPF:99 and NSOPF:04 and 32 replicate weights in NSOPF:93). These weights implement the balanced half-sample (BHS) method of variance estimation. They have been created to handle the certainty strata and to incorporate finite population correction factors for each of the noncertainty strata. Two widely available software packages, WesVar and PC CARP, have the capability to use replicate weights to estimate variances.

Analysts should be cautious about the use of BHS-estimated variances that relate to one stratum or to a group of two or three strata. Such variance estimates may be based upon far fewer than the number of replicates; thus, the variance of the variance estimator may be large. Analysts who use either the restricted-use faculty file or the institution file should also be cautious about cross-classifying data so deeply that the resulting estimates are based upon a very small number of observations. Analysts should interpret the accuracy of the NSOPF statistics in light of estimated standard errors and the small sample sizes.

Nonsampling Error
To minimize the potential for nonsampling errors, the NSOPF:04 Institution and Faculty questionnaires (as well as the sample design, data collection, and data processing procedures) were field-tested with a national probability sample of 150 postsecondary institutions (though only 80 of these were used for the full second-stage sampling of faculty and instructional staff) and 1,200 faculty members. A major focus of the field test was the effect of combining NSOPF and NPSAS. The field test also included an incentive experiment, which tested the use of incentives for increasing early responses and for obtaining interviews from nonrespondents. Other aspects of data quality were also examined.

The NSOPF:99 Institution and Faculty questionnaires (as well as the sample design, data collection, and data processing procedures) were field-tested with a national probability sample of 160 postsecondary institutions and 510 faculty members. Four methodological experiments—to increase unit response rates, speed the return of mail questionnaires, increase data quality, and improve the overall efficiency of the data collection process—were conducted as part of the field test. The experiments involved the use of prenotification, prioritized mail, a streamlined instrument, and the timing of CATI attempts. Another focus of the field test was the effort to reduce discrepancies between the faculty counts derived from the list of faculty provided by each institution and those provided in the Institution Questionnaire. Changes introduced to reduce discrepancies included providing clearer definitions of faculty eligibility (with consistency across forms and questionnaires) and

collecting list and Institution Questionnaire data simultaneously (with the objective of increasing the probability that both forms would be completed by the same individual and evidence fewer inconsistencies).

During the NSOPF:93 field test, a subsample of faculty respondents was reinterviewed to evaluate reliability. In addition, an extensive item nonresponse analysis of the field-tested questionnaires was conducted, followed by additional evaluation of the NSOPF:93 instruments and survey procedures. An item nonresponse analysis was also conducted for the full-scale data collection. Later, in 1996, NCES analyzed discrepancies in the NSOPF:03 faculty counts, conducting a retrieval, verification, and reconciliation effort to resolve problems.

Table 7. Summary of weighted response rates for selected NSOPF surveys

Questionnaire	List participation rate	Questionnaire response rate	Overall
NSOPF:93			
Institution	†	94	94
Faculty	84	83	70
NSOPF:99			
Institution	†	93	93
Faculty	88	83	74
NSOPF:04			
Institution	†	84	84
Faculty	91	76	69

†Not applicable.
SOURCE: Abraham, S.Y., Steiger, D.M., Montgomery, M., Kuhr, B.D., Tourangeau, R. Montgomery, B., and Chattopadhyay, M. (2002). *1999 National Study of Postsecondary Faculty* (NSOPF:99) *Methodology Report* (NCES 2002-154). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Huer, R., Kuhr, B., Fahimi, M., Curtin, T.R., Hinsdale, M., Carley-Baxter, L., and Green, P. (2005). *2004 National Study of Postsecondary Faculty (NSOPF:04) Methodology Report* (NCES 2006-179). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Selfa, L.A., Suter, N., Myers, S., Koch, S., Johnson, R.A., Zahs, D.A., Kuhr, B.D., and Abraham, S.Y. (1997). *1993 National Study of Postsecondary Faculty Methodology Report* (NCES 97-467). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Coverage Error. Because the IPEDS universe is the institutional frame for NSOPF, coverage of institutions is complete. However, there are concerns about the coverage of faculty and instructional staff. In NSOPF:04, prior to sampling, faculty counts from all lists provided by participating institutions were checked against both IPEDS and the counts that institutions provided in their Institution Questionnaire. (In NSOPF:99, the IPEDS comparison was used as a quality control check only when Institution Questionnaire counts were absent.) In NSOPF:04, as in NSOPF:99, institutions were contacted to resolve any discrepancies between data sources.

In NSOPF:99, in an effort to decrease the discrepancies in faculty counts noticed in NSOPF:93, ICs were asked to provide counts of full- and part-time faculty and instructional staff at their institutions as of November 1, 1998, the same reference date used for the 1997-98 IPEDS Fall Staff Survey; asked them to return both the faculty list and the Institution Questionnaire at the same time; and—giving them explicit warnings about potential undercounts of faculty—asked them to ensure that the counts provided in the list and questionnaire were consistent. These efforts appear to have worked, with 73 percent of institutions in NSOPF:99 providing questionnaire and list data that exhibited discrepancies of less than 10 percent, an improvement of 31 percentage points since NSOPF:93.

In NSOPF:93, a discrepancy between the faculty counts reported in the Institution Questionnaires and those provided in faculty lists by institutions at the beginning of the sampling process necessitated the "best estimates" correction to the NSOPF:93 faculty population estimates, as described earlier (in "Weighting," section 4).

Nonresponse Error.
Unit Nonresponse. Unit response rates have been similar over NSOPF administrations, though they decreased slightly in NSOPF:04 (see table 7). Note that the overall faculty response rates are the percentage of faculty responding in institutions that provided faculty lists for sampling.

Item Nonresponse. For the NSOPF:04 Institution Questionnaire, 2 of the 90 items had more than 15 percent of the data missing. For the Faculty Questionnaire, 34 of the 162 items had more than 15 percent of the data missing. For further details on item nonresponse, see the *2004 National Study of Postsecondary Faculty (NSOPF:04) Methodology Report* (Huer et al. 2005).

For the NSOPF:99 Institution Questionnaire, the mean item nonresponse rate was 3.4 percent (weighted). Overall, the item nonresponse rate for the Faculty Questionnaire was 6.2 percent. More than half of the items in the Faculty Questionnaire (55 percent) had an item nonresponse rate of less than 5 percent, 25 percent had rates between 5 and 10 percent, and 20 percent had rates greater than 10 percent. For further details on item nonresponse, see the *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (Abraham et al. 2002).

For the NSOPF:93 Institution Questionnaire, the mean item nonresponse rate was 10.1 percent, with the level of nonresponse increasing in the latter parts of the questionnaire. For the Faculty Questionnaire, the mean item nonresponse rate was 10.3 percent.

Measurement Error. In NSOPF:04, as in prior administrations of this study, secured faculty lists were evaluated for accuracy and completeness of information before being processed for sampling. To facilitate quality control, faculty list counts were compared against counts obtained from the following supplementary sources:

- the Institution Questionnaire (or the file layout form, if a questionnaire was not completed but an overall faculty count was supplied);
- the 2001 IPEDS Fall Staff Survey;
- the Contact Information and File Layout (CIFL) form (which included faculty counts and was used when questionnaire data was unavailable); and
- NSOPF:99 frame data.

Discrepancies in counts of full- and part-time faculty between the faculty list and other sources that were outside the expected range were investigated. All institutions with faculty lists that failed any checks were recontacted to resolve the observed discrepancies. Because of time and definitional differences between NSOPF and IPEDS, it was expected that the faculty counts obtained from the institutions and IPEDS would include discrepancies. Consequently, quality control checks against IPEDS were less stringent than those against the Institution Questionnaire. However, list count comparisons against IPEDS and NSOPF:99 data were useful in identifying systematic errors, particularly those related to miscoding of the employment status of faculty members.

Results of the data quality evaluations showed that 82 percent of faculty list counts were within 10 percent of the corresponding Institution Questionnaire counts. There were greater variances between list counts and IPEDS, which is based on a narrower definition of faculty. Patterns of discrepancies between IPEDS and list data followed expected patterns, with list counts larger than counts from IPEDS. For more information, see the *2004 National Study of Postsecondary Faculty (NSOPF:04) Methodology Report* (Huer et al. 2005).

For NSOPF:99, NCES conducted an intensive follow-up with 230 institutions (29 percent of those participating) whose reports exhibited a variance of 5 percent or more between the list and questionnaire counts overall or between the two part-time counts. NSOPF has experienced discrepancies in faculty counts among IPEDS, Institution Questionnaires, and faculty lists across all cycles of the study. Even though identical information is requested in the questionnaire and in the list (e.g., in NSOPF:99, a count of all full- and part-time faculty and instructional staff as of November 1, 1998), institutions have continued to provide discrepant faculty data. As in NSOPF:93, large discrepancies tend to be concentrated among smaller institutions and 2-year institutions in NSOPF:99. Undercounting of part-time faculty and instructional staff without faculty status in the list remains the primary reason for the majority of these discrepancies.

However, procedures implemented in NSOPF:99 improved the consistency of the list and questionnaire counts when compared to previous cycles of NSOPF. The percentage of institutions providing list and questionnaire data that had less than a 10 percent discrepancy increased from 42 percent in NSOPF:93 to 73 percent in NSOPF:99. A total of 43 percent provided identical data in the list and questionnaire in NSOPF:99 (compared to only 2.4 percent in NSOPF:93). Moreover, schools providing identical list and questionnaire data were shown to have provided more accurate and complete data in both the list and questionnaire. These findings suggest that the changed procedures that were introduced in the 1998 field test and NSOPF:99 resulted in more accurate counts of faculty and instructional staff. Institutions may also be in a better position to respond to these requests for data. Their accumulated experience in handling NSOPF and IPEDS (and other survey) requests, their adoption of better reporting systems, more flexible computing systems and staff, and a general willingness to provide the information are probably also a factor in their ability to provide more consistent faculty counts, although data to support these assertions are not available. For more detail, see the *1999 National Study*

of Postsecondary Faculty (NSOPF:99) Methodology Report (Abraham et al. 2002).

NCES conducted three studies to examine possible measurement errors in NSOPF:03, including (1) a reinterview study of selected faculty questionnaire items, conducted after the field test; (2) a discrepancy and trends analysis of faculty counts in the full-scale data collection; and (3) a retrieval, verification, and reconciliation effort involving recontact of institutions. For detail on these studies, see *Measurement Error Studies at the National Center for Education Statistics* (Salvucci et al. 1997) and the *1993 National Study of Postsecondary Faculty Methodology Report* (Selfa et al. 1997).

Reinterview Study. A reliability reinterview study was conducted after the NSOPF:93 field test to identify Faculty Questionnaire items that yielded low-quality data and the item characteristics that caused problems, thus providing a basis for revising the questionnaire items prior to implementation of the full-scale data collection. The analysis of the reinterview items was presented by item type—categorical or continuous variables—rather than by subject area. The level of consistency between the field-test responses and the reinterview responses was relatively high: a 70 percent consistency for most of the categorical variables and a 0.7 correlation for most of the continuous variables. A detailed analysis of the question on employment sector of last main job was conducted because it showed the highest percentage of inconsistent responses (28 percent) and the highest inconsistency index (36.0). It was concluded that the large number of response categories and the involvement of some faculty in more than one job sector were plausible reasons for the high inconsistency rate. The items with the lowest correlations were those asking for retrospective reporting of numbers that were small fractions of dollars or hours and those asking for summary statistics on activities that were likely to fluctuate over time—the types of questions shown to be unreliable in past studies.

Discrepancy and Trends Analysis of Faculty Counts. This analysis compared discrepancies between different types of institutions to identify systematic sources of discrepancies in faculty estimates between the list counts provided by the institutions and the counts they reported in the Institution Questionnaire. The investigation found that list estimates tended to exceed questionnaire estimates in large institutions, in institutions with medical components, and in private schools. Questionnaire estimates tended to be higher in smaller institutions, in institutions without medical components, and in public schools. Institutions supplied much higher questionnaire estimates than list estimates for part-time faculty. Faculty lists submitted early in the list collection process showed little difference in the magnitude of questionnaire/list discrepancies from faculty lists submitted later in the process.

Retrieval, Verification, and Reconciliation. This effort involved recontacting 509 institutions: 450 institutions (more than half of all institutions) whose questionnaire estimate of total faculty differed from their list estimate by 10 percent or more and an additional 59 institutions NCES designated as operating medical schools or hospitals. All institutions employing health sciences faculty and participating in NSOPF:93 were selected for recontact.

NCES accepted the reconciled estimates obtained in this study as the true number of faculty. More than half (57 percent) of the recontacted institutions identified the questionnaire estimate as the most accurate response, while 25 percent identified the list estimate as the most accurate. Another 11 percent of the institutions provided a new estimate; 1 percent indicated that their IPEDS estimate was the most accurate response; and 6 percent could not verify any of the estimates and thus accepted the original list estimate.

The majority of discrepancies in faculty counts resulted from the exclusion of some full- or part-time faculty from the list or questionnaire. Another factor was the time interval between the date the list was compiled and the date the questionnaire was completed. Downsizing also affected faculty counts at several institutions. Some of the reasons for the discrepancies were unexpected. For example, some institutions provided “full-time equivalents” (FTEs) on the Institution Questionnaire instead of an actual headcount of part-time faculty.

Sometimes part-time faculty were overreported—often as a result of confusion over the pool of part-time and temporary staff employed by, or available to, the institution during the course of the academic year versus the number actually employed during the fall semester. Another reason for overreporting part-time faculty was an inability to distinguish honorary/unpaid part-time faculty from paid faculty and teaching staff. This study also confirmed that a small number of institutions, those that considered their medical schools separate from their main campuses, excluded medical school faculty from their lists of faculty.

While these results indicate that there may have been some bias in the NSOPF:93 sample, no measure of the

potential bias, such as the net difference rate, was computed. Instead, the reconciliation prompted NCES to apply a poststratification adjustment to the estimates based entirely on the "best" estimates obtained during the reinterview study described above. Problems with health science estimates, however, could only be partly rectified by the creation of new "best" estimates. For more information on the calculation of the "best" estimates and further discussion of the health science estimates, see the *1993 National Study of Postsecondary Faculty Methodology Report* (Selfa et al. 1997).

Data Comparability

Design Changes. Each succeeding cycle of NSOPF has expanded the information base about faculty. NSOPF:04 was designed both to facilitate comparisons over time and to examine new faculty-related issues that had emerged since NSOPF:99. The NSOPF:04 sample was designed to allow detailed comparisons and high levels of precision at both the institution and faculty levels. The merging of NSOPF with NPSAS for the 2003–04 administration allowed for the inclusion of a larger number of institutions in NSOPF while reducing respondent burden. Since NSOPF:93, the operant definition of "faculty" for NSOPF has included instructional faculty, noninstructional faculty, and instructional personnel without faculty status.

NSOPF:04, NSOPF:99, and NSOPF:93 consisted of two questionnaires: an Institution Questionnaire and a Faculty Questionnaire. NSOPF:88 included, in addition, a Department Chairperson Questionnaire.

Definitional Differences. *Comparisons among the cycles must be made cautiously because the respondents in each cycle were different.* At the institution level, the NSOPF:04 sample consisted of all public and private, not-for-profit Title IV-participating, 2- and 4-year degree-granting institutions in the 50 states and the District of Columbia. The sample was first constituted in this way in NSOPF:99 so that the NSOPF sampling universe would conform with that of IPEDS. In the two previous rounds of the study (NSOPF:93 and NSOPF:88), the sample consisted of public and private, not-for-profit 2- and 4-year (and above) higher education institutions.

The definition of faculty and instructional staff for each NSOPF cycle is given above (see Section 3, "Key Concepts"). On the design level, note that NSOPF:04, NSOPF:99, and NSOPF:93 requested a listing of *all faculty (instructional and noninstructional) and instructional staff* from institutions for the purpose of sampling. For NSOPF:88, institutions were asked to provide only the names of *instructional faculty*. Although not specifically stated, NCES expected that institutions would provide information on instructional staff as well. The term faculty was used generically. However, there is no way of knowing how many institutions that had instructional staff as well as instructional faculty provided the names of both. Each institution was allowed to decide which faculty members belonged in the sample, thereby creating a situation that does not allow researchers to precisely match the *de facto* sample definition used by institutions in NSOPF:88.

Content Changes. Major goals for NSOPF:04 included making the questionnaires shorter and easier to complete. Other changes were implemented to bring NSOPF up to date with current issues in the field. As a result, 9 items from the NSOPF:99 Institution Questionnaire were eliminated from the NSOPF:04 Institution Questionnaire, 14 items were revised, and 3 items were repeated without change. For the NSOPF:04 Faculty Questionnaire, 39 items from the NSOPF:99 Faculty Questionnaire were eliminated, 51 items were simplified or otherwise revised, 1 item was added, and 3 items were unchanged.

Comparisons with other surveys. Comparisons of NSOPF:93 salary estimates with salary estimates from IPEDS and from the American Association of University Professors indicate that NSOPF data are consistent with these other sources. Most differences are relatively small and can be easily explained by methodological differences between the studies. The NSOPF estimates are based on self-reports of individuals, whereas the other two studies rely on institutional reports of salary means for the entire institution.

However, the reader should be aware of differences in faculty definitions between NSOPF and IPEDS. In IPEDS, individuals have to be categorized according to their primary responsibility (administrator, faculty, or other professional); in NSOPF, it is possible to categorize individuals according to any of their responsibilities.

Because NSOPF includes all faculty and instructional staff, it is possible for an "other professional" to have instructional responsibilities and/or be a faculty member, and it is also possible for an administrator to have instructional responsibilities and/or be a faculty member. Therefore, NSOPF includes all faculty under IPEDS, some of the administrators under IPEDS, and some of the other professionals under IPEDS.

6. CONTACT INFORMATION

For content information on NSOPF, contact:

Aurora M. D'Amico
Phone: (202) 502-7334
E-mail: aurora.damico@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Abraham, S.Y., Steiger, D.M., Montgomery, M., Kuhr, B.D., Tourangeau, R. Montgomery, B., and Chattopadhyay, M. (2002). *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (NCES 2002-154). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Abraham, S.Y., Suter, N.A., Spencer, B.D., Johnson, R.A., Zahs, D.A., and Myers, S.L. (1994). *1992–93 National Study of Postsecondary Faculty Field Test Report* (NCES 93-390). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Broyles, S.G. (1995). *Integrated Postsecondary Education Data System Glossary* (NCES 95-822). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Huer, R., Kuhr, B., Fahimi, M., Curtin, T.R., Hinsdale, M., Carley-Baxter, L., and Green, P. (2005). *2004 National Study of Postsecondary Faculty (NSOPF:04) Methodology Report* (NCES 2006-179). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Russell, S.H., Hancock, M.P., and Williamson, C. (1990). *1988 National Survey of Postsecondary Faculty Methodology Report*. Menlo Park, CA.

Selfa, L.A., Suter, N., Myers, S., Koch, S., Johnson, R.A., Zahs, D.A., Kuhr, B.D., and Abraham, S.Y. (1997). *1993 National Study of Postsecondary Faculty Methodology Report* (NCES 97-467). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Survey Design

Abraham, S.Y., Steiger, D.M., Tourangeau, R., Kuhr, B.D., Wells, B., and Yang, Y. (2000). *1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report* (NCES Working Paper 2000-01). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Wise, L.L., and McLaughlin, D.H. (1980). *Guidebook for Imputation of Missing Data* (SAGE). Palo Alto, CA: American Institutes for Research.

Data Quality and Comparability

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Chapter 14: National Postsecondary Student Aid Study (NPSAS)

1. OVERVIEW

SAMPLE SURVEY OF POSTSECONDARY INSTITUTIONS AND STUDENTS; CONDUCTED EVERY 3 to 4 YEARS

NPSAS collects information from:

- Student institutional record abstracts
- U.S. Department of Education administrative records
- Student interviews
- Parent interviews

The National Postsecondary Student Aid Study (NPSAS) is a comprehensive nationwide study conducted periodically by the National Center for Education Statistics (NCES) to determine how students and their families pay for postsecondary education. It is designed to address policy questions resulting from the rapid growth of financial aid programs and the succession of changes in financial aid program policies since 1986. The first NPSAS was conducted in the 1986–87 academic year (NPSAS:87). The seventh and most recently completed in the series was administered in the 2007–08 academic year (NPSAS:08). Other administrations have been conducted in academic year 1989–90 (NPSAS:90), 1992–93 (NPSAS:93), 1995–96 (NPSAS:96), 1999–2000 (NPSAS:2000), and 2003–04 (NPSAS:04).

NPSAS is based on a nationally representative sample of all students in eligible postsecondary education institutions in the 50 states, the District of Columbia, and Puerto Rico. Sampled institutions represent all major sectors, including public and private, not-for-profit and for-profit, and less-than-2-year schools, community colleges, 4-year colleges, and major universities with graduate-level programs. Study members include both undergraduate and graduate students who receive financial aid as well as those who do not. NPSAS data are obtained from administrative records of student financial aid, interviews with students, and, in prior cycles, interviews with a subsample of parents. Information has been gathered on as many as 130,000 students in a study cycle.

NPSAS also provides baseline data for two longitudinal studies: the Beginning Postsecondary Students Longitudinal Study (BPS) and the Baccalaureate and Beyond Longitudinal Study (B&B; see chapters 15 and 16, respectively). NPSAS:90, NPSAS:96, and NPSAS:04 served as baselines for BPS cohorts; NPSAS:93, NPSAS:2000, and NPSAS:08 were the baselines for B&B cohorts.

Unlike prior administrations, NPSAS:04 was conducted as the student component study of the 2004 National Study of Faculty and Students (NSoFaS:04). The faculty component—the 2004 National Study of Postsecondary Faculty (NSOPF:04)—was conducted primarily as a separate study, with the exception of institution sampling and contacting (see chapter 13). In both NPSAS NPSAS:04 and NPSAS:08 study samples were supplemented to provide representative estimates by institutional sector for several states.

Purpose

The purpose of the NPSAS is to produce reliable national estimates of characteristics related to financial aid for postsecondary students, the role of financial aid in how students and their families finance postsecondary education, and the extent to which the financial aid system is meeting the needs of students and families.

Components

NPSAS collects data on students from several sources, including: student records at the institution attended, student interviews, the Federal Student Aid Central Processing System (CPS), the National Student Loan Data System (NSLDS), the National Student Clearinghouse (NSC), ACT and SAT files, and the IPEDS Institutional Characteristics (IC) file.

Student Record Collection. The following information on students is obtained from institutional records: year in school, major field of study, type and control of institution, attendance status, tuition and fees, admission test scores, financial aid awards, cost of attendance, student budget information and expected family contribution for aided students, grade point average, age, and date first enrolled. Typically, an appointed Institutional Coordinator or a field data collector extracts the information from student records at a sample institution and enters it into a secure, customized web data collection system. In some cases, institutions and centralized systems choose to create and transmit a data file containing this information for all sample students from the sample institution(s).

Student Interview. Web-based student interviews (completed as a telephone interview or by self-administration) provide data on level (undergraduate, graduate, first-professional), major field of study, financial aid at other schools attended during the year, other sources of financial support, reasons for selecting the school currently being attended, current marital status, age, race/ethnicity, sex, highest degree expected, employment and income, voting in recent elections, and community service.

U.S. Department of Education Administrative Records. Since NPSAS:96, the following information has been collected from U.S. Department of Education Central Processing System (CPS) and National Student Loan Data System (NSLDS): types and amounts of federal financial aid received, cumulative Pell Grant and Stafford loan amounts, and loan repayment status. In NPSAS:08, information was also obtained for recipients of the new Academic Competitiveness Grant (ACG) and the National Science and Mathematics Access to Retain Talent Grant (National SMART Grant).

Other administrative databases. Data collected from commercial databases, such as: enrollment, degree, and certificate records from the National Student Clearinghouse (NSC); and ACT and SAT test score. data

Parent Interview. Telephone interviews with a limited sample of students' parents (conducted through NPSAS:96) collected supplemental data, including parents' marital status, age, highest level of education achieved, income, amount of financial support provided to children, types of financing used to pay children's educational expenses, and occupation and industry.

Out-of-School Student Loan Recipient Survey. This survey was only conducted as part of NPSAS:87. It collected data on major field of study; years attended and degrees received (if any); type and control of institution; financial aid; aid repayment status; age; sex; race/ethnicity; marital status; income; and employment history (occupation, industry, and salary).

Periodicity

Triennial from 1986–87 through 1995–96, and quadrennially beginning in 1999–2000. The next data collection is scheduled for 2012.

2. USES OF DATA

The goal of the NPSAS study is to identify institutional, student, and family characteristics related to participation in financial aid programs. Federal policymakers use NPSAS data to determine future federal policy concerning student financial aid. With these data, it is possible to analyze special population enrollments in postsecondary education, including students with disabilities, racial and ethnic minorities, students taking remedial/developmental courses, students from families with low incomes, and older students. The distribution of students by major field of study can also be examined. Fields of particular interest are mathematics, science, and engineering, as well as teacher preparation and health studies. Data can also be generated on factors associated with choice of postsecondary institution, participation in postsecondary vocational education, parental support for postsecondary education, and occupational and educational aspirations.

It is important that statistical analyses be conducted using software that properly accounts for the complex sampling design of NPSAS. NCES has recently developed new software tools for analysis of complex survey data: QuickStats allows users to generate simple tables and graphs quickly, and PowerStats allows researchers to generate more complex tables and run linear and logistic regressions. Data from NPSAS:04 and NPSAS:08 can be analyzed with QuickStats and PowerStats. The Data Analysis System (DAS) may be used for analyses using NPSAS data prior to 2003-04. For information on other software packages and

statistical strategies useful for analysis of complex survey data, see appendix M of *the 2004 National Postsecondary Student Aid Study (NPSAS:04) Full-Scale Methodology Report* (Cominole et al. 2006).

3. KEY CONCEPTS

Described below are several key concepts relevant to financial assistance for postsecondary education. For additional NPSAS terms, refer to the glossaries in published statistical analysis reports and database documentation.

Institution Type. A derived variable that combines information on the level and control of the NPSAS institution. Institution level concerns the institution's length of program and highest degree offering and is defined as less than 2-year, 2- to 3-year, 4-year nondoctorate, or 4-year doctorate (including first-professional degree). Institution control concerns the source of revenue and control of operations and is defined as public, private not-for-profit, or private for-profit.

Attendance Pattern. A student's intensity and persistence of attendance during the NPSAS year. Intensity refers to whether the student attended full- or part-time while enrolled. Persistence refers to the number of months a student is enrolled during the year. Students are considered to be enrolled for a full year if they are enrolled 8 or more months during the year. Months do not have to be contiguous or at the same institution, and students do not have to be enrolled for a full month to be considered enrolled for that month. In surveys prior to NPSAS:96, a full year was defined as 9 or more months.

Dependency Status. If a student is considered financially dependent, the parents' assets and income are considered in determining aid eligibility. If the student is financially independent, only the student's assets are considered, regardless of the relationship between student and parent. The federal definition of dependency status has remained the same in each administration of NPSAS from academic year 1995–96 through 2007–08. All students who are age 24 or over in the fall term of the NPSAS year are considered to be independent. Students under 24 who are married, have legal dependents other than a spouse, are veterans, or are an orphan or ward of the courts are also independent. Other undergraduates under age 24 are considered to be dependent, unless they can demonstrate to a financial aid officer that they do not receive any financial support from their parents. All graduate and professional students in programs beyond a bachelor's degree are considered to be independent.

Expected Family Contribution (EFC). The amount of financial support for the student's undergraduate education that is expected to be provided by the student's family, or directly by the student if the student is financially independent. This amount is used to determine financial need and is based upon dependency status (see above definition), family income and assets, family size, and the number of children in the family enrolled in postsecondary education. This information is gathered from the Department of Education's financial aid system (the Central Processing System), or it is imputed from student income.

Title IV Financial Aid. The sum of the following types of federal aid: Pell Grants, Supplemental Educational Opportunity Grants (SEOG), Perkins Loans, Stafford Loans, PLUS Loans, and Federal Work Study. NPSAS:08 also included Academic Competitiveness Grants and National SMART Grants.

4. SURVEY DESIGN

Target Population

The *target population* is defined as all eligible students enrolled at any time during the federal financial aid award year in postsecondary institutions in the United States or Puerto Rico that have a signed Title IV participation agreement with the U.S. Department of Education (thus making these institutions eligible for federal student aid programs). The population includes both students who receive aid and those who do not receive aid. It excludes students who are enrolled solely in a general equivalency diploma (GED) program or are concurrently enrolled in high school.

Sample Design

The design for the NPSAS sample involves the selection of a nationally representative sample of postsecondary education institutions and students within these institutions. Prior to NPSAS:96, a geographic-area-clustered, three-stage sampling design was used to: (1) construct geographic areas from three-digit postal zip code areas; (2) sample institutions within the geographic sample areas; and (3) sample students within sample institutions. Beginning with NPSAS:96, the sample design eliminated the first stage of sampling (geographic area construction), thereby increasing the precision of the estimates. Institutional and student sample sizes vary somewhat from cycle to cycle depending on study design and budget

considerations at the time. Approximately 1,960 institutions and 137,800 students were initially selected for participation in NPSAS:08.

Institution Sample. To be eligible for inclusion in the institution sample, an institution must satisfy the following conditions: (1) offer an education program designed for persons who have completed secondary education; (2) offer an academic, occupational, or vocational program of study lasting at least 3 months or 300 clock hours; (3) offer access to the general public; (4) offer more than just correspondence courses; (5) be located in the 50 states, the District of Columbia, or Puerto Rico; and (6) be other than a U.S. Service Academy. Also, beginning with NPSAS:2000, eligible institutions must have a signed Title IV participation agreement with the U.S. Department of Education.

The institution-level sampling frame is constructed from the Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) and header files (see chapter 12). Although the institutional sampling strata have varied across NPSAS administrations, in all years the strata are formed by classifying institutions according to control (public or private), level, and highest degree offering. The NPSAS:04 strata were also formed by Carnegie classification and state, and the NPSAS:08 strata were also formed by state. A stratified sample of institutions is then selected with probability proportional to size. School enrollment, as reported in the IPEDS, defines the measure of size; enrollment is imputed if missing in the IPEDS file. Institutions with expected frequencies of selection greater than unity are selected with certainty. The remainder of the institution sample is selected from the other institutions within each stratum.

Additional implicit stratification is accomplished within each institutional stratum by sorting the stratum sampling frame in a serpentine manner. Implicit stratification allows the approximation of proportional representation of institutions on additional measures. In NPSAS:08, the implicit strata were formed using (1) Historically Black Colleges and Universities (HBCU) indicator; (2) Hispanic-Serving Institutions (HSI) indicator; (3) Carnegie classifications of postsecondary institutions; (4) the Office of Business Economics (OBE) Region from the IPEDS header file (Bureau of Economic Analysis of the U.S. Department of Commerce Region); and (5) an institution measure of size. Further implicit stratification was done for the State Univerisity of New York (SUNY) and City University of New York (CUNY) systems in New York, the state and technical colleges in Georgia, and the state universities in California.

In NPSAS:04, the implicit strata were formed using (1) the HBCU indicator; (2) Carnegie classifications (3) OBE Region; and (4) an institution measure of size. In NPSAS:2000, for less-than-2-year, 2-year, and private for-profit institutions, the implicit strata were formed using (1) institutional level of offering (where levels had been collapsed to form strata); (2) the OBE Region from the IPEDS header file; (3) the Federal Information Processing Standard (FIPS) state code; and (4) an institution measure of size. For public 4-year and private not-for-profit 4-year institutions, the implicit strata were formed using (1) Carnegie classifications of institutions or groupings of Carnegie classifications; (2) the HBCU indicator; (3) the OBE Region from the IPEDS header file; and (4) an institution measure of size. In NPSAS:96, the implicit strata were formed using (1) institutional level of offering; (2) the IPEDS IC-listed U.S. Department of Commerce Region; and (3) an institution measure of size. Selected institutions are asked to verify their IPEDS classification (institutional control and highest level of offering) and the calendar system that they use (including dates that terms start).

The NPSAS:08 institution sampling frame was constructed from the 2004–05 IPEDS IC, header, and Fall Enrollment files and, because NPSAS:08 also serves as the base-year survey for a longitudinal cohort of baccalaureate recipients (i.e., B&B), the 2004–05 IPEDS Completions file. A total of 1,960 of the 6,780 institutions in the survey universe were selected for the NPSAS:08 sample. The sampled institutions were stratified into 22 national strata and 24 state strata based on institutional control, institutional offering, and highest degree offering.

The institutional sampling frame for NPSAS:04 was constructed from the 2000–01 IPEDS IC, header, and Fall Enrollment files; 1,670 of the 7,710 institutions in the survey universe were selected for NPSAS:04. The sampled institutions were stratified into 22 national strata and 36 state strata based on institutional control, institutional offering, highest degree offering, and Carnegie classification. The institutional sampling frame for NPSAS:2000 was constructed from the 1998–99 IPEDS IC file and, because NPSAS:2000 also served as the base-year survey for a B&B cohort, the 1996–97 IPEDS Completions file. Eligible institutions were partitioned into 22 institutional strata based on institutional control, highest level of offering, and percentage of baccalaureate degrees awarded in education. Approximately 1,100 institutions were initially selected for NPSAS:2000. As noted above, NPSAS:96 was the first administration of NPSAS to employ a single-stage institutional sampling design, no longer constructing geographic areas as the initial step.

The sampling frame for NPSAS:96 was the 1993–94 IPEDS IC file; 9,470 of the 10,650 institutions in the file were deemed eligible for NPSAS:96. Eligible institutions were stratified into nine strata based on institutional control and highest level of offering.

Student Sample. Full- and part-time students enrolled in academic or vocational courses or programs at eligible institutions, and not concurrently enrolled in a high school completion program, are eligible for inclusion in NPSAS. NPSAS:87 sampled students enrolled in the fall of 1986. Beginning with NPSAS:90, students enrolled at any time during the year were eligible for the study. This design change provided the data necessary to estimate full-year financial aid awards.

Sampled institutions are asked to provide student enrollment lists with the following information for each student: full name, identification number, Social Security number, educational level, an indication of first-time beginning student (FTB) status or baccalaureate recipiency (depending on the longitudinal cohort being launched), major, and, beginning with NPSAS:04, a local address, a local telephone number, a campus e-mail, a permanent address, a permanent phone number, and a permanent e-mail. Additionally, date of birth and class level of undergraduates were requested for NPSAS:08. The student sample is drawn from these lists, which were provided by 1,730 of 1,940 eligible institutions in NPSAS:08; 1,360 of 1,630 eligible institutions in NPSAS:04; 1,000 of the nearly 1,100 eligible institutions in NPSAS:2000; and 840 of 900 eligible institutions in NPSAS:96.

Basic student sample. Students are sampled on a flow basis (using stratified systematic sampling) from the lists provided by institutions. Steps are taken to eliminate both within- and cross-institution duplication of students. NPSAS classifies students by educational level as undergraduate, master's, doctor's, other graduate, or first-professional students. For the purpose of defining the third cohort of B&B, NPSAS:08 classified undergraduates into (1) business major potential baccalaureate recipients, (2) other potential baccalaureate recipients, and (3) other undergraduates. Potential baccalaureate recipients were further stratified by those who are science, technology, engineering, or mathematics (STEM) majors and all other majors and by SMART Grant recipients and non-recipients. Other undergraduates were further stratified by SMART Grant recipients, Academic Competitiveness Grant (ACG) recipients, and non-recipients. The categories for potential baccalaureate recipients and other undergraduates were then stratified by in-state and out-of-state status. NPSAS:04 stratified undergraduate students as (1) potential FTBs and (2) other undergraduates. These two categories were then stratified by in-state and out-of-state status. The FTBs in NPSAS:04 make up the third cohort of BPS. For the purpose of defining the second cohort of B&B, NPSAS:2000 also broke down undergraduates into: (1) business major baccalaureate recipients, (2) other baccalaureate recipients, and (3) other undergraduates. In NPSAS:96, FTBs, or students beginning their postsecondary education during one of the terms of the NPSAS:96 sample year composed the second cohort of the BPS, with the data collected serving as the base-year data for the subsequent longitudinal studies.

The student sample is allocated to the combined institutional and student strata (e.g., graduate students in public 4-year doctorate institutions). Initial student sampling rates are calculated for each sample institution using refined overall rates to approximate equal probabilities of selection within the institution-by-student sampling strata. These rates are sometimes modified to ensure that the desired student sample sizes are achieved.

In NPSAS:08, adjustments to the initial sampling rates resulted in some additional variability in the student sampling rates and, hence, in a likely increase in survey design effects. Such rate adjustment procedures have generally proven effective. The overall sample yield in NPSAS:08 was close to expected (137,800 students vs. the target of 138,000). The student sample consisted of 29,470 potential baccalaureate recipients; 95,650 other undergraduates; 6,530 master's students; 3,760 doctoral students; 470 other graduate students; and 1,920 first-professional students.

Initial sampling rates were adjusted in NPSAS:04, NPSAS:2000, and NPSAS:96, as well. The overall sample yield in NPSAS:04 was less than expected (109,210 students vs. the target of 121,680). The student sample consisted of 49,410 FTBs; 47,680 other undergraduates; 3,720 master's students; 4,950 doctoral students; 1,660 other graduate students; and 1,790 first-professional students. (See "FTB sample" below for more detail on the sampling of FTBs.) In NPSAS:2000, the overall sample yield was very close to expected (70,230 students vs. the target of 70,270). The student sample consisted of 57,600 undergraduates; 5,960 master's students; 3,950 doctoral students; 1,370 other graduate students; and 1,350 first-professional students. In NPSAS:96, the overall sample yield was actually greater than expected (63,620 students vs. the target of 59,510). The student sample consisted of 23,610 potential FTBs; 27,540

other undergraduates; 9,690 graduate students; and 2,780 first-professional students.

Student interview sample. NPSAS:04 was the first administration of NPSAS to offer the option of self-administration of the student interview via the Web, in addition to computer-assisted telephone interviewing (CATI). In NPSAS:08, these procedures resulted in 95,360 completed interviews, about two-thirds of which were completed by self-administration and one-third by CATI. In NPSAS:04, these procedures resulted in 62,220 completed interviews, 28,710 of which were completed by self-administration and 33,510 by CATI.

In NPSAS:2000, student interviews were conducted primarily by CATI. To help reduce the level of nonresponse to CATI, computer-assisted personal interviewing (CAPI) procedures, using field interviewers, were used for the first time. Of the 66, 340 eligible students in the initial CATI sample, some 51,010 were located for CATI interviewing, while 11, 960 were "unlocatable" in CATI and were eligible for field locating and/or CAPI; the rest were either ineligible or excluded.

Due to budget limitations, NPSAS:96 attempted CATI interviews for only a subsample of the basic student sample. A two-phase, nonrespondent follow-up subsampling design was used to maximize the yield of completed student interviews obtained from the CATI subsample while achieving acceptable response rates. These procedures resulted in 51,200 students being selected for Phase 1 of the CATI interviewing. A sample of nonrespondents to Phase 1 was selected for Phase 2 with specified rates based on the outcome of the Phase 1 efforts and the seven sampling strata; 25,770 students were selected for Phase 2.

Parent interview subsample. In NPSAS:96, a subsample of students selected for the student interview was also designated for parent interviews. In the Phase 1 CATI subsample of NPSAS:96, students were designated for parent interviews if they met one of the following criteria: they were dependent undergraduate students not receiving federal aid; they were dependent undergraduate students receiving federal aid whose parents' adjusted gross income was not available; or they were independent undergraduate students who were 24 or 25 years old on December 31, 1995. All 8,800 students who fell into one of these groups were sampled for parent interviews. The parent interview was discontinued after NPSAS:96.

Longitudinal Study Samples. In NPSAS:90, a new longitudinal component collected baseline data for students who started their postsecondary education in the 1989–90 academic year. These students were followed over time in BPS, with the first follow-up in 1992. Beginning postsecondary students from NPSAS:96 and NPSAS:04 were also followed up and surveyed two and five years later. Similarly, NPSAS:93, NPSAS:2000, and NPSAS:08 provided baseline data for students who received baccalaureates in the 1992–93, 1999–2000, and 2007–08 academic years, respectively. These graduates have been followed over time as part of B&B. The next cohort of BPS will be identified in NPSAS:12 and follow-up studies will be conducted in 2014 and 2017.

BPS sample. Final FTB status is determined based on data from several sources: enrollment lists, student record data, student interviews, loan history data from NSLDS, and enrollment history data from NSC.

First, however, institutions are asked to identify potential FTBs in the student lists they provide. However, the information available to institutions is often insufficient for determining an accurate count of FTBs; for example, students transferring from another institution without transfer credits might mistakenly be counted as FTBs. In NPSAS:04, FTB sampling rates were based primarily on the BPS experience in NPSAS:96, which indicated that the number of students listed as potential FTBs who were not actual FTBs far exceeded the number of students not identified as potential FTBs who later proved to be FTBs. As in the past, the NPSAS:04 longitudinal cohort was oversampled to support the next round of BPS.

B&B sample. B&B:08 is the third cohort in the B&B series and the second to gather college transcript data on such a longitudinal sample. The first B&B longitudinal cohort was identified in NPSAS:93 and consisted of students who received their bachelor's degree in academic year 1992–93. NPSAS:93 provided the base-year data, and students were interviewed in an initial follow-up in 1994; this follow-up also included a collection of transcript data. The 1993 cohort was surveyed again in 1997 and 2003. The first transcript collection was conducted as part of B&B:93/94. The second B&B cohort was selected from NPSAS:2000, which became the base year for a single follow-up in spring 2001.

The B&B:08 sample consists of students eligible to participate in the NPSAS:08 full-scale study who completed requirements for the bachelor's degree in the 2007–08 academic year. The first follow-up study (B&B:08/09) involved two data collection components. First, postsecondary transcripts were collected from each of the NPSAS institutions where sample members

completed their program requirements. It was followed by an interview focusing on plans after degree completion.

B&B status is determined on the basis of multiple sources: student enrollment lists from institutions, student record collection, student interviewing, and transcripts (in B&B:93/94 and B&B:08/09).

Data Collection and Processing

Reference Dates. Data are collected for the financial aid award year, which spans from July 1 of one year through June 30 of the following year.

Data Collection. NPSAS involves a multistage effort to collect information related to student aid. The first stage involves collecting applicants from the U.S. Department of Education's Central Processing System (CPS).

Another stage of data collection involves collecting information from the student's records at the school from which he or she was sampled. Since NPSAS:93, these data have been collected through a computerized system, which facilitates both the collection and transfer of information to subsequent electronic systems. To reduce respondent burden, several data elements are preloaded into the records collection system records prior to collection at the institution. These include student demographics, Student Aid Report information on federal financial aid applicants, and nonfederal aid common to a particular institution. Institutional Coordinators are given the option of having their staff or contractor field data collectors perform the data collections. About 66 percent of the institutions in NPSAS:04, as well as 74 percent in NPSAS:2000, and 57 percent in NPSAS:96 chose self-administration, using a computer-based program to provide student record data. In NPSAS:08, very few institutions (about 1 percent) chose the field interviewer option for completion. Approximately 63 percent chose self- administration, and 36 percent provided the student record data via electronic files (primarily large institutions or systems).

In the student interview stage of data collection, information on family characteristics, demographic characteristics, and educational and work experiences and aspirations is obtained from students. Student and parent paper questionnaires were used to collect this information in NPSAS:87, but beginning with NPSAS:90, student and parent data were collected by computer-assisted-telephone-interviewing (CATI). Parent interviews, however, were not conducted after NPSAS:96. NPSAS:04 was the first administration of NPSAS to offer students the opportunity to participate by self-administered web surveys or by CATI, an approach that continued in NPSAS:08.

The NPSAS:08 student interview contained six sections and was programmed for both self-administered web surveys and CATI. An abbreviated interview was developed that contained a subset of key items from the main interview. This version was used during refusal conversion toward the end of data collection. The abbreviated interview was also translated into Spanish for telephone administration to Spanish speakers with limited English proficiency.

The student interview included an online coding system used to obtain IPEDS information for postsecondary institutions (other than the NPSAS institution from which the student was sampled) that the student attended during the same year. After the respondent or interviewer provided the state and city in which the institution is located, the online coding system displayed the list of all postsecondary institutions in that location, and the respondent/interviewer could select the appropriate institution. Upon selection, the name of the institution, as well as selected IPEDS variables (institutional level, control), was inserted into the database.

An assisted coding system was also developed to facilitate the coding of major/field of study into categories that can be mapped to values in NCES's Classification of Instructional Programs (CIP).

The data collection design for student interviewers has evolved over time. In NPSAS:2000, student interviews were conducted primarily by telephone, and occasionally in person, using CATI/CAPI technology. In NPSAS:04 and NPSAS:08 abbreviated interviews were developed to convert refusals toward the end of data collection, and an online coding system was used, to obtain IPEDS information. NPSAS:96 differed from other cycles in that only a subsample of the initial student sample was selected for the interview stage (in order to reduce overall costs for the study).

The final stage of data collection involves retrieval of additional Student Aid Report (ISAR) data (for the academic year beyond the NPSAS year) from the Central Processing System (CPS), data on Pell Grant applications for the NPSAS year from the Pell Grant file, and data on recipients of Academic Competitiveness Grants and SMART Grant, as well as loan histories of applicants for federal student loans from the National Student Loan Data System (NSLDS). All of these files are maintained by the U.S. Department of Education. Additional data for the NPSAS sample are obtained from other sources as

well, including test score data from the ACT and College Board (SAT), enrollment data from the National Student Clearinghouse (NSC), and data from the Veterans Administration.

Editing. Initial editing takes place during data entry. The web-based data collection systems used for the student interview and student record collection have built-in quality control checks to notify users of invalid or out-of-range entries. For example, the student records collection system will notify the user of any student records that are incomplete (and the area of incompleteness) and any records that have not yet been accessed. A pop-up screen provides full and partial completion rates for institutional record collection. Data are subjected to edit checks for completeness of critical items.

Following the completion of data collection, all student record and interview data are edited to ensure adherence to range and consistency checks. Range checks are summarized in the variable descriptions contained in the data files. Inconsistencies, either between or within data sources, are resolved in the construction of derived variables. Items are checked for validity by comparing the student interview responses to information available in institutional records. Missing data codes characterize blank fields as don't know/data not available; refused; legitimate skip; data source not available (not applicable to the student); or other.

Estimation Methods

Weighting is used to adjust NPSAS data to national population totals and to adjust for unit nonresponse. Imputation is used to compensate for item nonresponse and mitigate associated bias.

Weighting. For the purpose of obtaining nationally representative estimates, sample weights are created for both the institution and the student. Additional weighting adjustments, including nonresponse and poststratification adjustments, compensate for potential nonresponse bias and frame errors (differences between the survey population and the ideal target population). The weights are also adjusted for multiplicity at the institution and student levels and for unknown student eligibility.

In NPSAS:08 and NPSAS:04, the institution weight was computed first and then used as a component of the student weight. Student weights were calculated as the product of the total of 10 weight components for NPSAS:08 and 13 weight components for NPSAS:04, each representing either a probability of selection or a weight adjustment.

In NPSAS:2000, statistical analysis weights were computed for two sets of respondents: CATI respondents and other study respondents. These were calculated as the product of 13 weight components, again representing either a probability of selection or a weight adjustment.

In NPSAS:96, study weights were applied to students who responded to specified student record or CATI data items. Study and CATI weights were calculated as the product of 14 weight components. First-time beginning students (FTBs) whose first postsecondary institution was not the NPSAS sample institution were not included in BPS. To compensate for their exclusion, FTB weights were computed by making a final weighting class adjustment to the CATI weights by institution type. All adjustment factors were close to one, ranging from 1.00 to 1.02. The development of the student record weight components was similar to the development of the study and CATI weight components—except that the student record components applied to a different set of respondent data and did not include the CATI weight components.

Imputation. When the editing process (including logical imputations) is complete, the remaining missing values for all variables with missing data are statistically imputed in order to reduce the bias of survey estimates caused by missing data. Variables are imputed using a weighted sequential hot-deck procedure whereby missing data are replaced with valid data from donor records that match the recipients with respect to the matching criteria.

In NPSAS:08 and NPSAS:04, variables requiring imputation were not imputed simultaneously. However, some variables that were related substantively were grouped together into blocks, and the variables within a block were imputed simultaneously. Basic demographic variables were imputed first using variables with full information to determine the matching criteria. The order in which variables were imputed was also determined to some extent by the substantive nature of the variables. For example, basic demographics (such as age) were imputed first and these were used to process education variables (such as student level and enrollment intensity), which, in turn, were used to impute financial aid variables (such as aid receipt and loan amounts).

For variables with less than 5 percent missing data, the variables used for matching criteria were selected based on prior knowledge about the dataset and the known relationships between the variables. For variables with more than 5 percent missing data, a

statistical process called Chi-Squared Automatic Interaction Detection (CHAID) was used to identify the matching criteria that were most closely related to the variables being imputed.

In NPSAS:2000, the remaining missing values for 23 analysis variables were imputed statistically; most of the variables were imputed using a weighted hot-deck procedure. To implement the weighted hot-deck procedure, imputation classes and sorting variables relevant to each item being imputed were defined. If more than one sorting variable was chosen, a serpentine sort was performed where the direction of the sort (ascending or descending) changed each time the value of a variable changed. The serpentine sort minimized the change in the student characteristics every time one of the variables changed its value.

The respondent data for five of the items being imputed were modeled using a CHAID analysis to determine the imputation classes. These items were parent income (imputed for dependent students only), student income (imputed for independent students only), student marital status, local residence, and a dependents indicator.

A CHAID analysis was performed on these variables because of their importance to the study and the large number of candidate variables available with which to form imputation classes. Also, for the income variables, trying to define the best possible imputation classes was important due to the large amount of missing data. The CHAID analysis divided the respondent data for each of these five items into segments that differed with respect to the item being imputed. The segmentation process first divided the data into groups based on categories of the most significant predictor of the item being imputed. It then split each of these groups into smaller subgroups based on other predictor variables. It also merged categories of a variable that were found insignificant. This splitting and merging process continued until no more statistically significant predictors were found (or until some other stopping rule was met). The imputation classes were then defined from the final CHAID segments.

In NPSAS:96, some 22 analysis variables were statistically imputed. All variables, with the exception of the estimated family contribution were imputed using a weighted hot-deck procedure. First, the respondent data for six key items were modeled using a CHAID analysis to determine the imputation classes. These items were race/ethnicity, parent income (for dependent students only), student income, student marital status, a dependents indicator, and number of dependents. Then, 21 items imputed by the weighted hot-deck approach. The remaining 15 items were: parent family size, parent marital status, student citizenship, student gender, student age, dependency status, local residence, type of high school degree, high school graduation year, fall enrollment indicator, attendance intensity in fall term, student level in last term, student level in first term, degree program in last term, and degree program in first term. Only four of these 15 items had more than 5 percent of their cases imputed: parent family size (18 percent), parent marital status (16 percent), high school degree (5 percent), and high school graduation year (5 percent).

Recent Changes

NPSAS:04 included important new features in sample design and data collection. For the 2004 study, NPSAS and NSOPF were conducted together under one contract: the 2004 National Study of Faculty and Students (NSoFaS:04). There has historically been a great deal of overlap in the institution samples for these two studies since the target populations for both involve postsecondary institutions. To minimize institutional burden, and to maximize efficiency in data collection procedures, the two studies were combined.

Another important change in NPSAS:04 was that it was designed to provide state-level representative estimates for undergraduate students within three institutional strata—public 2-year institutions, public 4-year institutions, and private not-for-profit 4-year institutions—in 12 states that were categorized into three groups based on population size (four large, four medium, and four small): California, Connecticut, Delaware, Georgia, Illinois, Indiana, Minnesota, Nebraska, New York, Oregon, Tennessee, and Texas. NPSAS:08 was designed to provide state-level representative estimates for undergraduates within four institutional strata—public 2-year institutions, public 4-year institutions, private not-for-profit 4-year institutions, and private for-profit degree-granting 2-year-or-more institutions. In NPSAS:08, state-level estimates were provided for California, Texas, New York, Illinois, Georgia, and Minnesota.

Also of importance is the inclusion of an option for self-administration via the Web of the student interview in NPSAS:04. This option was provided in addition to CATI interviews, which were employed in past rounds of NPSAS. Regardless of completion mode, a single web-based instrument was employed.

NPSAS:08 was again conducted independently of the NSOPF study but carried along all of the technical innovations and design enhancemements of prior rounds. It was also designed to provide state-level

representative estimates for undergraduates within four institutional strata—public 2-year institutions, public 4-year institutions, private not-for-profit 4-year institutions, and private for-profit degree-granting 2-year-or-more institutions. In NPSAS:08, state-level estimates were provided for California, Texas, New York, Illinois, Georgia, and Minnesota.

The most significant enhancement to NPSAS:2000 involved the development and implementation of a new web-based system for use in the student record abstraction process. This web-based software had an improved user interface compared to the NPSAS:96 system and addressed several of the student records collection issues raised during NPSAS:96 (e.g., insufficient computer memory, failures during diskette installation and virus scanning, and lack of information regarding institutions' progress during data collection).

Other changes in NPSAS:2000 included: adding a series of questions about financial aid, as a new way of obtaining information about financial assistance received from sources other than federal student aid; adding several new items intended to capture the increased use of technology among students; and adding a new eligibility requirement for postsecondary institutions—to have a signed Title IV participation agreement with the U.S. Department of Education during the NPSAS academic year.

NPSAS:96 introduced important new features in sample design and data collection. It was the first NPSAS to employ a single-stage institutional sampling design (no longer using an initial sample of geographic areas and institutions within geographic areas). This design change increased the precision of study estimates. NPSAS:96 was also the only NPSAS to select a subsample of students for telephone interviews and to take full advantage of administrative data files. Through file matching/downloading arrangements with the Department of Education's Central Processing System, the study obtained financial data on federal aid applicants for both the NPSAS year and the following year. Through similar arrangements with the National Student Loan Data System, full loan histories were obtained. Cost efficiencies were introduced through a dynamic two-phase sampling of students for CATI, and the quality of collected institutional data was improved through an enhanced student records collection procedure. New procedures were also introduced to broaden the base of postsecondary student types for whom telephone interview data could be collected: the use of Telephone Display for the Deaf technology to facilitate telephone communications with hearing-impaired students, and a separate Spanish translation interview for administration to students with limited English language proficiency.

Future Plans

The next NPSAS data collection (NPSAS:12) is scheduled for 2012 and will serve as the base for the fourth cohort of BPS (BPS:12/14, BPS:12/17).

5. DATA QUALITY AND COMPARABILITY

Every major component of the study is evaluated on an ongoing basis so that necessary changes can be made and assessed prior to task completion. Separate training is provided for CADE and CATI data collectors, and interviewers are monitored during CATI operations for deviations from item wording and skipping of questions. The CATI system includes online coding of postsecondary education institution and major field of study, so that interviewers can request clarification or additional information at the time of the interview. Quality circle meetings of interviewers, monitors, and supervisors provide a forum to address work quality, identify problems, and share ideas for improving operations and study outcomes. Even with such efforts, however, NPSAS—like every survey—is subject to various types of errors, as described below.

Sampling Error

Because NPSAS samples are probability-based samples rather than simple random samples, simple random sample techniques for estimating sampling error cannot be applied to these data. Two procedures for estimating variances, the Taylor Series linearization procedure and the Jackknife replicate procedure, are available for use with NPSAS:96 data. The Taylor Series linearization procedure and the balanced repeated replication (BRR) procedure are available on the NPSAS:2000 data files. The Taylor Series linearization procedure and the bootstrap replication procedure are available on the NPSAS:08 and NPSAS:04 data files.

Taylor Series. For NPSAS:96, analysis strata and replicates for three separate datasets were defined: all students, all undergraduate students, and all graduate/first-professional students. For NPSAS:2000, analysis strata and replicates for four separate datasets were defined: all students, all undergraduate students, all graduate/first-professional students, and all baccalaureate recipients. For NPSAS:08 and NPSAS:04, analysis strata and replicates were defined for the combined set of all students.

Jackknife. In NPSAS:96, the Jackknife analysis strata were defined to be the same as the analysis strata defined for the Taylor Series procedure. Based on the

Jackknife strata and replicate definitions, seven replicate weight sets were created—one set for the CADE weights and three sets each for the study and CATI weights. The study and CATI sets included separate replicate weights for all students, undergraduates only, and graduates only.

Balanced Repeated Replication. The BRR procedure is an alternative variance estimation procedure that computes the variance based on a balanced set of pseudo-replicates. To form pseudo-replicates for BRR variance estimation, the Taylor Series analysis strata were collapsed. The number of Taylor Series analysis strata and primary sampling units were different for all students combined, graduates/first-professionals, and baccalaureate recipients, so the collapsing was done independently and, hence, with different results. Replicate weights were created, associated with the two analysis weights: study weights and CATI weights. Thus, a total of five replicate weight sets were created for NPSAS:2000. For the study weights, this included separate replicate weights for all students and for graduate/first-professional students only; for the CATI weights, this included separate replicate weights for all students, graduate/first-professional students only, and baccalaureates only.

Bootstrap. In NPSAS:08 and NPSAS:04, a vector of bootstrap sample weights was added to the analysis file to facilitate computation of standard errors for both linear and nonlinear statistics. These weights are zero for units not selected in a particular bootstrap sample; weights for other units are inflated for the bootstrap subsampling. The initial analytic weights for the complete sample are also included for the purpose of computing the desired estimates. The vector of replicate weights allows for computing additional estimates for the sole purpose of estimating a variance. The replicate in NPSAS:08 were produced using methodology adapted from Kott (1998) and Flyer (1987) and those in NPSAS:04 weights were produced using a methodology and computer software developed by Kaufman (2004). NPSAS:08 included 200 replicate weights.

Nonsampling Error

Coverage Error. Because the institutional sampling frame is constructed from the IPEDS IC file, there is nearly complete coverage of the institutions in the target population. Student coverage, however, is dependent upon the enrollment lists provided by the institutions. In NPSAS:08, approximately 1,730 of the 1,940 eligible institutions provided student lists or databases that could be used for sample selection. A total of 1,360 of the 1,630 eligible institutions in NPSAS:04; 1,000 of the nearly 1,100 eligible institutions in NPSAS:2000; and 840 of the 900 eligible institutions in NPSAS:96 provided student lists or databases that could be used for sample selection.

Several checks for quality and completeness of student lists are made prior to actual student sampling. In NPSAS:96 and NPSAS:04, *completeness checks* failed if (1) FTBs were not identified (unless the institution explicitly indicated that no such students existed) or (2) student level (undergraduate, graduate, or first professional) was not clearly identified. In NPSAS:2000 and NPSAS:08, completeness checks failed if (1) baccalaureate recipients/graduating seniors were not identified, (2) student level was not clearly identified, or (3) major fields of study or CIP codes were not clearly identified for baccalaureates.

Quality checks are performed by comparing the unduplicated counts (by student level) in institution lists with the nonimputed unduplicated counts in IPEDS IC files. Institutions failing these checks were called to rectify the problems before sampling began. These checks were performed through the 2007–08 administration. In NPSAS:08, after any necessary revisions, all but seven lists submitted were usable for selecting the student sample; in NPSAS:04, all but two lists submitted were usable for selecting the student sample.

Nonresponse Error. The response rates described in this section refer to NPSAS:08.

Unit nonresponse. Some 90 percent (weighted) of eligible sample institutions provided student enrollment lists. The total weighted student response rate was 96 percent. Table 8 provides a summary of response rates across NPSAS administrations.

There are several types of participation/coverage rates in NPSAS. In NPSAS:08, institution participation rates were generally lowest among for-profit institutions and institutions whose highest offering is less than a 4-year program.

For the *student record abstraction* phase of the study (referred to as CADE), institution completion rates were 94 percent (weighted) for institutions choosing field-CADE, 96 percent for institutions choosing self-CADE, and 98 percent for data-CADE (submitting data via electronic files). CADE completion rates varied by type of institution, ranging from 92 percent for private not-for-profit less-than-2-year institutions to 100 percent for private not-for-profit less-than-4-year institutions. Overall, the student-level CADE completion rate (the percentage of NPSAS-eligible sample members for whom a completed CADE record

was obtained) was 96 percent (weighted). Weighted student-level completion rates ranged from 87 percent for private, not-for-profit, less-than-4-year institutions to 99 percent for public, 4-year, non-doctorate-granting institutions. Weighted completion rates by student type were 97 percent for undergraduate and 98 percent for graduate and first-professional students.

Table 8. Weighted response rates for selected NPSAS administrations.

Component	Institution list participation rate	Student response rate	Overall
NPSAS:96			
Student survey (analysis file[1])	91	96	88
Student survey (student interview)	91	76	70
NPSAS:2000			
Student survey (analysis file[1])	91	97	89
Student survey (student interview)	91	72	66
NPSAS:04			
Student survey (analysis file[1])	80	91	72
Student survey (student interview)	80	71	56
NPSAS:08			
Student survey (analysis file[1])	90	96	86
Student survey (student interview)	90	71	64

—Not available.
[1]NPSAS analysis file contains analytic variables derived from all NPSAS data sources (including institutional records and extant data sources) as well as selected direct student interview variables.
NOTE: The student interview response rates for NPSAS:96 and NPSAS:2000 are for CATI interviews only. The response rates for student interviews in NPSAS:04 include all interview modes.
SOURCE: Cominole, M.B., Siegel, P.H., Dudley, K., Roe, D., and Gilligan, T. (2006). *2004 National Postsecondary Student Aid Study (NPSAS:04) Full-Scale Methodology Report* (NCES 2006-180). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Riccobono, J.A., Cominole, M.B., Siegel, P.H., Gabel, T.J., Link, M.W., and Berkner L.K. (2001). *National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000) Methodology Report* (NCES 2002-152). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Riccobono, J.A., Whitmore, R.W., Gabel, T.J., Traccarella, M.A., Pratt, D.J., and Berkner, L.K. (1997). *National Postsecondary Student Aid Study, 1995–96 (NPSAS:96) Methodology Report* (NCES 98-073). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Wei, C.C., Berkner, L., He, S., Lew, S., Cominole, M., and Siegel, P. (2009). *2007–08 National Postsecondary Student Aid Study (NPSAS:08): Student Financial Aid Estimates for 2007–08: First Look* (NCES 2009-166). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Overall, 95,360 of approximately 132,800 eligible sample members (72 percent unweighted) completed either a full or partial NPSAS:08 student interview. The weighted response rate was 71 percent overall and ranged from 56 percent for private, for-profit, less-than-2 year institutions to 77 percent for public, 4-year, doctorate-granting institutions.

Item nonresponse. Each NPSAS institution is unique in the type of data it maintains for its students. Because not all desired information is available at every institution, the CADE software allows entry of a "data not available" code. In NPSAS:08, item response rates student record abstraction were very high overall. Two items had low response rates: marital status (46 percent) and additional phone numbers (17 percent). Thus, student records frequently lack these items. The other items had response rates ranging from 73 percent to just below 100 percent.

Missing data for items in the NPSAS:08 student interview were associated with several factors: (1) a true refusal to answer, (2) an unknown answer, (3) confusion over the question wording or response options, or (4) hesitation to provide a "best guess" response. Item nonresponse rates were based on the

number of interview respondents to whom the item was applicable and of whom it was asked. Overall, item-level nonresponse rates were low, with only 23 items out of approximately 500 having more than 10 percent of data missing

Measurement Error. Due to the complex design of NPSAS, there are several possible sources of measurement error, as described below.

Sources of response. Each source of information in NPSAS has both advantages and disadvantages. While students are more likely than institutions to have a comprehensive picture of education financing, they may not remember or have records of exact amounts and sources. This information may be more accurate in student financial aid records and government databases since it is recorded at the time of application for aid.

Institutional records. While financial aid offices maintain accurate records of certain types of financial aid provided at their own institution, these records are not necessarily inclusive of all support and assistance. They may not maintain records of financial aid provided at other institutions attended by the student, and they may not include employee educational benefits and institutional assistantships, which are often treated as employee salaries. These amounts are assumed to be underreported.

Government databases. Federal aid information can only be extracted from federal financial aid databases if the institution can provide a valid Social Security number for the student. It is likely that there is some undercoverage of federal aid data in NPSAS.

CATI question delivery and data entry. Any deviation from item wording that changes the intent of the question or obscures the question meaning can result in misinterpretation on the part of the interviewee and an inaccurate response. CATI entry error occurs when the response to a question is recorded incorrectly. Measures of question delivery and data entry are used for quality assurance monitoring. Due to ongoing monitoring of student telephone interviews, problems are usually detected early and the CATI interviewers are retrained, if necessary. Overall error rates in NPSAS:08 were low (typically below 2 percent) and within control limits.

Self-administered web survey. Self-administration introduces challenges not experienced with single-mode interviewer-administered surveys. For instance, in self-administration, interviewers are not able to clarify question intent and probe when responses are unclear. Surveys also require modifications to account for the mixed-mode presentation (i.e., self-administered and CATI) to maintain data quality and to make the interview process as efficient as possible for respondents. These considerations were addressed in the design of the survey, making the two modes as consistent as possible.

Data Comparability

As noted above, important design changes have been implemented in NPSAS across administrations. While sufficient comparability in survey design and instrument was maintained to ensure that comparisons with past NPSAS studies could be made, data from the later studies are not comparable to data from the first study (NPSAS:87) for the following reasons: (1) NPSAS:87 only sampled students enrolled in fall 1986, whereas the later studies sampled from enrollments covering a full year; and (2) NPSAS:87 did not include students from Puerto Rico, whereas NPSAS:90 and later studies have included a small sample of Puerto Rican students. However, users of NPSAS data files can produce estimates for the later studies comparable to those from NPSAS:87 by selecting only students enrolled in the fall and excluding those sampled from Puerto Rico. Note also that the method used to generate the lists of students from which to sample was changed for NPSAS:93 and later NPSAS studies.

Comparisons with IPEDS Data. NCES recommends that readers not try to produce their own estimates (e.g., the percentage of all students receiving aid or the numbers of undergraduates enrolled in the fall who receive federal or state aid) by combining estimates from NPSAS publications with IPEDS enrollment data. The IPEDS enrollment data are for fall enrollment only and include some students not eligible for NPSAS (e.g., those enrolled in U.S. Service Academies and those taking college courses while enrolled in high school).

6. CONTACT INFORMATION

For content information on NPSAS, contact:

Aurora M. D’Amico
Phone: (202) 502-7334
E-mail: aurora.damico@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Berkner, L.K., He, S., Lew, S., Cominole, M.B., and Siegel, P.H. (2005). *2003–04 National Postsecondary Student Aid Study (NPSAS:04) Student Financial Aid Estimates for 2003–04* (NCES 2005-158). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Cominole, M.B., Siegel, P.H., Dudley, K., Roe, D., and Gilligan, T. (2006). *2004 National Postsecondary Student Aid Study (NPSAS:04) Full-Scale Methodology Report* (NCES 2006-180). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Flyer, P.A. (1987). Finite Population Correction for Replication Estimates of Variance. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Alexandria, VA: American Statistical Association.

Kaufman, S. (2004). Using the Bootstrap in a Two-Stage Design When Some Second-Stage Strata Have Only One Unit Allocated. *Proceedings of the American Statistical Association, Section on Survey Research Methods*.

Kott, P.S. (1988). Model-Based Finite Population Correction for the Horvitz-Thompson Estimator. *Biometrika, 75*(4): 797–799.

Loft, J.D., Riccobono, J.A., Whitmore, R.W., Fitzgerald, R.A., and Berkner, L.K. (1995). *Methodology Report for the National Postsecondary Student Aid Study, 1992–93* (NCES 95-211). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Malitz, G. (1990). *Methodology Report for the National Postsecondary Student Aid Study, 1987* (NCES 90-309). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Malizio, A.G. (1992). *Methodology Report for the 1990 National Postsecondary Student Aid Study* (NCES 92-080). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Malizio, A.G. (2001). *National Postsecondary Student Aid Study: Student Financial Aid Estimates for 1999–2000* (NCES 2001-209). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Riccobono, J.A., Cominole, M.B., Siegel, P.H., Gabel, T.J., Link, M.W., and Berkner, L.K. (2001). *National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000) Methodology Report* (NCES 2002-152). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Riccobono, J.A., Whitmore, R.W., Gabel, T.J., Traccarella, M.A., Pratt, D.J., and Berkner, L.K. (1997). *National Postsecondary Student Aid Study, 1995–96 (NPSAS:96) Methodology Report* (NCES 98-073). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Wei, C.C., Berkner, L., He, S., Lew, S., Cominole, M., and Siegel, P. (2009). *2007–08 National Postsecondary Student Aid Study (NPSAS:08): Student Financial Aid Estimates for 2007–08: First Look* (NCES 2009-166). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Survey Design

Malizio, A.G. (1996). *National Postsecondary Student Aid Study: 1996 Field Test Methodology Report* (NCES Working Paper 96-17). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Data Quality and Comparability

Ault, K., Black, S., Chromy, J., Fahimi, M., Siegel, P.H., Trofimovich, L., Whitmore, R.W., and Berkner, L.K. (2003). *Imputation Methodology for the National Postsecondary Student Aid Study: 2004* (NCES Working Paper 2003-20). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Griffith, J. and Hunt-White, T. (2009) *2007-08 National Postsecondary Student Aid Study (NPSAS:08) Restricted-Use Data File* (NCES 2010-171). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Vavricka, J. and Nevill, S. (2009) 2007-08 National Postsecondary Student Aid Study (NPSAS:08): Undergraduate Data Analysis System (NCES 2009-167). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 15: Beginning Postsecondary Students (BPS) Longitudinal Study

1. OVERVIEW

The Beginning Postsecondary Students (BPS) Longitudinal Study was implemented in 1990 to complement the NCES longitudinal studies of high school cohorts and improve data on participants in postsecondary education. BPS draws its cohorts from the National Postsecondary Student Aid Study (NPSAS), which regularly collects financial aid and other data on nationally representative cross-sectional samples of postsecondary students (see chapter 14). NPSAS provides the base-year data for first-time beginning (FTB) postsecondary students; BPS then follows these students through school and into the workforce.

BPS includes nontraditional (older) students as well as traditional students and is, therefore, representative of *all* beginning students in postsecondary education. By starting with a cohort that has already entered postsecondary education and following it every 2 to 3 years for at least 6 years, BPS can describe to what extent, if any, students who start their education later differ in progress, persistence, and attainment from students who start earlier. In addition to the student data, BPS collects federal financial aid records covering the entire undergraduate period, providing complete information on progress and persistence in school.

The first BPS cohort followed a subset of NPSAS:90 respondents who began their postsecondary education in the 1989-90 academic year. About 8,000 eligible students from NPSAS:90 were included in the first and the second BPS follow-ups in 1992 and 1994. The second BPS cohort was based on NPSAS:96 with the first BPS follow-up in 1998 and the second in 2001. This cohort followed about 10,200 eligible students who started their postsecondary education in the 1995–96 academic year. The third BPS cohort was selected from NPSAS:04 and included students who began their postsecondary education in 2003–04. Approximately 18,600 students were determined to be eligible for inclusion in the third BPS cohort. The first follow-up with these students occurred in 2006 and the second in 2009.

LONGITUDINAL SAMPLE SURVEY OF FIRST-TIME BEGINNING POSTSECONDARY STUDENTS, INCLUDING BOTH TRADITIONAL AND NONTRADITIONAL STUDENTS

BPS includes:

- Base-year NPSAS data
- Student interviews
- Financial aid records

Purpose
To collect data related to persistence in and completion of postsecondary education programs, relationships between work and education, and the effect of postsecondary education on the lives of individuals.

Components
BPS consists of base-year data obtained from NPSAS, follow-up data collected in BPS surveys, student aid data from the U.S. Department of Education, including information from the Federal Student Aid Central Processing System (CPS), the National Student Loan Data System (NSLDS), and program data files, such as Pell and other grant programs; and administrative records available from the other sources (e.g., the National Student Clearinghouse).

Base-Year Data (from NPSAS). Base-year data for BPS are collected in NPSAS from students, parents (in the first and second cohorts only), institutional records, and Department of Education financial aid records. These data cover major field of study; type and control of institution; financial aid; cost of attendance; age; sex; race/ethnicity; family income; reasons for school selection; current marital status; employment and income; community service; background and preparation for college; college experience; future expectations; and parents' level of education, income, and occupation. These data represent the 1989–90 academic year for the first BPS cohort, the 1995–96 academic year for the second BPS cohort, and the 2003–04 academic year for the third BPS cohort.

BPS Follow-up Surveys. Follow-up data are obtained from student interviews and financial aid records on year in school; persistence in enrollment; academic progress; degree attainment; change in field of study; institution transfer; education-related experiences; current family status; expenses and financial aid; employment and income; employment-related training; community service; political participation; and future expectations.

Follow-ups for the first BPS cohort were conducted in spring 1992 (BPS:90/92) and spring 1994 (BPS:90/94). BPS:90/92 focused on continued education and experience, employment and financing, educational aspirations, and family formation. The focus of BPS:90/94 was on continuing education experiences and financing, including degree attainment and graduate/professional school access; employment experiences; educational and employment aspirations; and family formation.

The second BPS cohort participated in two follow-up surveys as well. These follow-ups were conducted in 1998 (BPS:96/98) and 2001 (BPS:96/01). The BPS:96/98 interview collected information on postsecondary enrollment, employment, income, family formation/household composition, student financial aid, debts, education experiences, and education and career aspirations. BPS:96/01 focused exclusively on activities since the BPS:96/98 interview, collecting information on postsecondary enrollment and degree attainment; undergraduate education experiences; postbaccalaureate education experiences (for those sample members who had completed a bachelor's degree since the last interview); employment; and family, financial, and disability status as well as civic participation since the last interview.

Follow-ups for the third BPS cohort were conducted in 2006 (BPS:04/06) and 2009 (BPS:04/09). The 2006 follow-up focused primarily on continued education and experience, education financing, entry into the workforce, the relationship between experiences during postsecondary education and various societal and personal outcomes, and returns to the individual and to society on the investment in postsecondary education. The second follow-up in 2009 focused primarily on employment, baccalaureate degree completion, graduate and professional school access issues, and returns to the individual and to society from the completion of a postsecondary degree. In addition, postsecondary transcripts were collected from all institutions attended by members of the third BPS cohort.

Periodicity

BPS cohorts are followed at least twice after first entering postsecondary education (as determined in NPSAS). Follow-ups take place at 2- to 3-year intervals.

2. USES OF DATA

BPS addresses persistence, progress, and attainment after entry into postsecondary education and also directly addresses issues concerning entry into the workforce. Its unique contribution is the inclusion of students who are not direct entrants to postsecondary education from high school, a steadily growing segment of the postsecondary student population. Their inclusion allows analysis of the differences, if any, between traditional (recent high school graduates) and nontraditional students in aspirations, progress, persistence, and attainment.

Congress and other policymakers use BPS data when they consider how new legislation will affect college students and others in postsecondary education. BPS data can answer such questions as: What percentage of beginning students complete their degree programs? What are the financial, family, and school-related factors that prevent students from completing their programs, and what can be done to help them? Do students receiving financial aid do as well as those who do not? Would it be better if the amount of financial aid was increased? Additional questions that BPS can address include the following: Do students who are part-time or discontinuous attenders have the same educational goals as full-time, consistent attenders? Are they as likely to attain similar educational goals? Are students who change majors more or less likely to persist?

3. KEY CONCEPTS

Institution Type. Defined by level of degree offering and length of program at the postsecondary institution. Institutions are generally classified as (1) less than 2-year (offers only programs of study that are less than 2 years in duration); (2) 2- to 3-year, sometimes referred to in reports as 2-year (confers at least a 2-year formal award, but not a baccalaureate degree, or offers a 2- or 3-year program that partially fulfills the requirements for a baccalaureate or higher degree at a 4-year institution; this category includes most community and junior colleges); and (3) 4-year (confers at least a baccalaureate degree and may also confer higher level degrees, such as master's, doctoral, and first-professional degrees; this category is often broken down into doctorate-granting vs. nondoctorate-granting).

Institution Control. Control of postsecondary institution is classified as follows: (1) public; (2) private not-for-profit; and (3) private for-profit.

First-Time Beginning Students (FTBs). The target population for BPS. For the first BPS cohort, FTBs were defined as students who enrolled in postsecondary education for the first time after high school in the 1989–90 academic year *(pure FTBs)*. Individuals who started postsecondary education earlier, left, and then returned were not included. The second BPS cohort comprised both students who enrolled for the very first time in the 1995–96 academic year and students who had previously enrolled but had not completed a postsecondary course for credit prior to July 1, 1995 *(effective FTBs)*. This expanded definition shifted the requirement from the act of enrollment to successful completion of a postsecondary course. The third BPS cohort comprised both students who enrolled for the first time in the 2003–04 academic year and those who had previously enrolled but had not completed a postsecondary course for credit prior to July 1, 2003.

Nontraditional Students. Primarily older students who delayed postsecondary enrollment; that is, students who did not enter postsecondary education in the same calendar year as high school graduation or who received a general equivalency diploma (GED) or other certificate of high school completion.

Persistence. Continuous enrollment in postsecondary education with the goal of obtaining a degree or other formal award.

Attainment. Receipt of the degree or other formal award while enrolled in postsecondary institutions.

4. SURVEY DESIGN

Target Population

All students who first entered postsecondary education after high school in the 1989–90 academic year (the first BPS cohort), the 1995–96 academic year (the second BPS cohort), and the 2003–04 academic year (the third BPS cohort). The definition of FTB was refined for the second and third BPS cohorts to include students who had enrolled in postsecondary education prior to completion of high school if they had not completed a postsecondary course for credit before July 1, 1995 (the beginning of the 1995–96 academic year, for the second BPS cohort) or July 1, 2003 (the beginning of the 2003–04 academic year, for the third BPS cohort). BPS includes students in nearly all types of postsecondary education institutions located in the 50 states, the District of Columbia, and Puerto Rico: public, private not-for-profit, and private for-profit institutions; 2-year, 2- to 3-year, and 4-year institutions; and occupational programs that last for less than 2 years. Excluded are students attending U.S. Service Academies, institutions that offer only correspondence courses, and institutions that enroll only their own employees. Generally BPS data are nationally representative by institutional level and control (for more information, readers should consult each study's methodological report). Data; the data are not representative at the state level.

Sample Design

Student eligibility for BPS is determined in two stages. The first stage involves selection for the base-year NPSAS sample; see chapter 14 for a description of NPSAS sample design and determination of FTBs who make up the BPS cohorts. All FTBs who complete interviews in NPSAS are considered eligible for BPS. The second stage involves a review of NPSAS data to see if any potential FTBs have been misclassified. FTB status for additional students may be determined through (1) reports from NPSAS institutions; (2) responses of the sample member during the BPS interview; and (3) modeling procedures used following data collection.

First BPS Cohort (1989–90). The first BPS cohort initially consisted of 11,700 students (from about 1,090 institutions) who had been interviewed in the 1989–90 NPSAS. In the second follow-up of this cohort (in 1994), a working sample of 7,910 individuals was initially used. It consisted of the first follow-up eligible respondents plus those nonrespondents for whom FTB status had yet to be determined. Only 7,130 sample members could be located. Of these, 6,790 members were interviewed (either fully or partially). Some of

those interviewed (170) were determined to be non-FTBs, leaving 6,620 eligible FTBs who were either fully (5,930) or partially (690) interviewed in the second follow-up.

Second BPS Cohort (1995–96). In the second BPS cohort, 12,410 confirmed and potential FTBs were selected (from about 800 institutions) for continued follow-up from a total NPSAS pool of 15,730 confirmed or potential FTBs. This pool included 3,740 who had not been interviewed in the 1995–96 NPSAS (of whom 430 were selected for potential continued inclusion in BPS). This BPS-eligible sample of 12,410 individuals was further reduced when an additional 230 were determined to be ineligible. The BPS-eligible sample contained 10,270 FTBs who were given full or partial interviews in the first follow-up; 1,060 were not able to be contacted, and 850 did not respond.

The final sample for this cohort included 10,370 individuals. This included all respondents to earlier follow-ups as well as a subsample of earlier nonrespondents and other individuals who were unavailable for earlier data collections.

Third BPS Cohort (2003–04). The third BPS cohort consisted of 23,090 confirmed and potential FTBs (from 1,360 institutions), of whom approximately 18,640 were determined to be eligible. Of this final BPS-eligible sample, 14,900 FTBs were given full or partial interviews in the first follow-up; 2,060 were not located, and 1,670 did not respond. Prior to the second follow-up, 30 sample members were determined to be deceased. Of the remaining sample, 15,160 FTBs were given full or partial interviews; 1,690 were not located; 1,440 did not respond; and 320 were determined to be exclusions.

Data Collection and Processing

For the first and second BPS cohorts, computer-assisted telephone interviewing (CATI) was the primary data collection tool. All locating, interviewing, and data processing activities were under the control of an Integrated Control System (ICS), consisting of a series of PC-based, fully linked modules. The various modules of the ICS provided the means to conduct, control, coordinate, and monitor the several complex, interrelated activities required in the study and served as a centralized, easily accessible repository for project data and documents.

For the third BPS cohort, a self-administered web interview was introduced as an additional data collection option. A single web-based instrument was developed for these self-administered interviews as well as for CATI interviews and computer-assisted personal interviews (CAPI). All aspects of the study for the third cohort were controlled using an Integrated Management System (IMS): a comprehensive set of desktop tools that included a management module, a Receipt Control System (RCS) module, and an instrumentation module.

BPS is conducted for NCES by the Research Triangle Institute.

The following sections describe the data collection and processing procedures for BPS follow-ups. Refer to chapter 14 for a description of data collection and processing for the base-year data obtained from NPSAS.

Reference Dates. The base-year (NPSAS) survey largely refers to experiences in postsecondary schooling in the academic year covered by NPSAS. The follow-ups cover the 2- to 3-year interval since the previous round of data collection. Some data are collected retrospectively for the previous survey.

Data Collection. Data collection in BPS follow-ups involves concerted mail and telephone efforts to trace potential sample members to their current location and to conduct a CATI interview both to establish study eligibility and collect data. Field locating and CAPI interviews were also used with the second and third cohorts. The third cohort introduced self-administered web interviews as an additional initial data collection method.

Locating students begins with information provided by the BPS locating database, which is updated by a national change-of-address service before the locating effort begins. Cases not located during the previous round of the survey are forwarded to pre-CATI telephone tracing and, subsequently, to field locating if intensive telephone tracing is unsuccessful. Prior to the start of CATI operations, a pre-notification mailing is sent to the student, and the current contact information is provided to interviewers for basic CATI locating. (For the third BPS cohort, there was an additional 4-week early response period during which sample members could complete a self-administered web interview before CATI operations began.) In the event that CATI locating is unsuccessful, cases are sent to post-CATI central telephone tracing and, again as necessary, field locating. During tracing operations, cases of “exclusion” are identified, such as those who are (1) outside of the calling area; (2) deceased; (3) institutionalized or physically/mentally incapacitated and unable to respond to the survey; or (4) otherwise unavailable for the entire data collection period.

Throughout the data collection period, interviewers are monitored for delivery of questionnaire text and recognition statements, probing, feedback, and CATI entry errors.

Each coding operation is subjected to quality control review and recoding procedures by expert coders. Subsequent to data collection, all “other, specify” responses are evaluated for possible manual recoding into existing categories or into new categories created to accommodate responses of high frequency through a process known as “upcoding.” Efforts are also made to convert several items with high rates of undetermined responses (including refusal or “don’t know”). In order to reduce indeterminacy rates for personal, parent, and household income items, as well as for other financial amount items, specific questions are included in the survey to route initial “don’t know” responses through a series of screens that seek closer and closer financial estimates.

In the second follow-up of the first BPS cohort, amount ranges for the “don’t know” conversion screens were based on frequencies obtained from the second follow-up field test for the same items. Indeterminacy conversion was attempted for five financial amount items (financial aid amount, total loan amount, respondent gross income, parents’ gross income, and household gross income) and was very successful for initial “don’t know” responses. Conversion rates were greater than 50 percent for every item attempted, with an overall success rate of 65 percent.

With the second BPS cohort, approximately 1,930 sample members initially refused to participate in the first follow-up. Fifty-three percent (1,020) of these refusals were converted. For the second follow-up of this cohort, 1,860 sample members refused to participate at least once. Of these, 74 percent were converted.

For the first follow-up of the third BPS cohort, 1,850 sample members refused to be interviewed at some point in the data collection. Of these refusals, 700 (approximately 38 percent) ultimately completed an interview. In the second follow-up, 8,380 sample members reached the nonrespondent phase of interviewing, with 4,860 (almost 58 percent) completing the interview before the end of data collection.

Editing. The CATI data are edited and cleaned as part of the preparation of the data file. Modifications to the data are made, to the extent possible, based on problem sheets submitted by interviewers, which detail item corrections, deletions, and prior omissions. In addition, variables are checked for legitimate ranges and interim consistency. Coding corrections and school information from the Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics files are merged into the CATI files. Data inconsistencies identified during analyses are also corrected, as appropriate and feasible.

In addition, the web instrument for the follow-up interviews with the third BPS cohort (BPS:04/06 and BPS:04/09) included online coding systems which ensured that most codes were assigned during data collection rather than during data editing. Post-data collection, data were edited using procedures developed for previous NCES-sponsored studies, including the base-year study (NPSAS:04). These included quality checks and examinations of skip patterns and the reasons for missing data.

Estimation Methods

Weighting is used to adjust for unit nonresponse. Only minimal imputation is performed to compensate for item nonresponse.

Weighting. BPS follow-ups involve further identification of FTB status for sample members who were in the earlier round of BPS. Furthermore, post hoc modeling is implemented following the first follow-up data collection in an attempt to identify non-FTBs among nonrespondents.

Four sets of weights were computed for use with BPS data for the first (1989–90) cohort: (1) 1992 cross-sectional weights for cross-sectional analyses of the first cohort at the time of the first follow-up, based on the first follow-up data collection; (2) 1994 cross-sectional weights for cross-sectional analyses of the first cohort at the time of the second follow-up data collection; (3) 1992 cross-sectional weights for the first follow-up information that was collected either during the first follow-up or retrospectively in the second follow-up; and (4) longitudinal weights for comparison of the responses pertaining to the 1990, 1992, and 1994 cross-sectional populations (e.g., trend analyses) for those students who responded to each of the three surveys: the 1989–90 NPSAS, the BPS first follow-up (in 1992), and the BPS second follow-up (in 1994). For computation of these weights, see the technical report for the second follow-up (BPS:90/94; Pratt et al. 1996).

The 1994 cross-sectional weights can also be used for longitudinal analyses involving data items collected retrospectively in the second follow-up, because those data items are available for 1992 (either directly from the first follow-up or retrospectively from the second follow-up if the student responded in 1994). Each set

of weights consists of an analysis weight for computing point estimates of population parameters, plus a set of 35 replicate weights for computation of sampling variances using the Jackknife replication method of variance estimation. All weight adjustments were implemented independently for each set of replicate weights. (See “Sampling Error” in section 5 below for further detail on replicate variance estimation.)

For the second BPS cohort, four sets of weights were also constructed: (1) 1998 analysis weights for point estimates of population parameters for students in the first follow-up (BPS:96/98); (2) 2001 cross-sectional weights for analyzing respondents to the second follow-up (BPS:96/01); (3) longitudinal weights for analyzing respondents to NPSAS:96 and both BPS follow-ups; and (4) longitudinal weights for analyzing respondents only to NPSAS:96 and BPS:96/01.

Analysis weights were also developed for the first follow-up of the third BPS cohort (BPS:04/06). These weights were derived from the NPSAS:04 weights. Three types of weights were developed for the analysis of data from the second follow-up (BPS:04/09): (1) cross-sectional weights for cases that were BPS:04/09 study respondents (i.e., had data from either the student interview or enrollment data from other external sources, (2) longitudinal or panel weights for cases who were study respondents for NSPAS:04, BPS:04/06, and BPS:04/09, and (3) weights for analyzing BPS:04/09 sample members with any transcript data.

Imputation. Imputation was performed on a small number of variables for the earlier cohorts of BPS. These variables relate to the student’s dependency status and family income in each survey round. For example, the variable containing dependency status for aid in academic year 1989–90 was derived by examining all applicable variables used in the federal definition of dependency for the purpose of applying for financial aid. If information was not available for all variables, dependency status was imputed based on age, marital status, and graduate enrollment. Similarly, the variable containing the 1988 family adjusted gross income used imputed values if responses were not available.

In the follow-ups for the second BPS cohort, logical imputations were performed where items were missing but their values could be implicitly determined, such as the amount earned by a respondent who did not work in 2000 (imputed to $0).

With the third BPS cohort, imputation was performed for all variables on the data file with missing data, including questionnaire items and derived variables. In addition, nonrespondents to the BPS:04/06 interview appear in the analysis file with imputed data. Response rates and estimated bias in BPS:04/06 are reported both with nonimputed data (prior to item imputation) and after imputation. For BPS:04/09, imputation was also performed for questionnaire items with missing data, including cases who did not complete the interview but had enrollment data from other sources.

Future Plans

The fourth BPS cohort (representing the 2011–12 academic year) will be selected in 2012 from the NPSAS:12 sample after the study’s student interview has been completed.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Because the NPSAS sample design involves stratification, disproportionate sampling of certain strata, and clustered (i.e., multistage) probability sampling, the standard errors, design effects, and related percentage distributions for a number of key variables in BPS have been calculated with the software package SUDAAN. These variables include sex, race/ethnicity, age in the base year, socioeconomic status, income/dependency in the base year, number of risk factors in the base year, level and control of the first institution, and aid package at the first institution in the base year. These estimates provide an approximate characterization of the precision with which BPS survey statistics can be estimated.

Several specific procedures are available for calculating precise estimates of sampling errors for complex samples. Taylor Series approximations, Jackknife repeated replications, and balanced repeated replications produce similar results.

Nonsampling Error

Nonsampling error in BPS is largely related to nonresponse bias caused by unit and item nonresponse and to measurement error.

Coverage Error. The BPS sample is drawn from NPSAS. Consequently, any coverage error in the NPSAS sample will be reflected in BPS. (Refer to chapter 14 for coverage issues in NPSAS.)

Nonresponse Error. Unit nonresponse is reported in BPS in terms of *contact rates* (the proportion of sample members who were located for an interview) and

interview rates (the proportion of sample members who fully or partially completed the interview). Item nonresponse has not been fully evaluated, although the numbers of nonrespondents are in the electronic codebook on an item-by-item basis.

Unit Nonresponse. The results for the second follow-up of the first BPS cohort (BPS:90/94) show a *contact rate* of 92 percent. The rate was substantially lower for individuals who did not respond to the first follow-up (75 percent) than for those who did respond (95 percent). Contact rates also varied by institution type. The rate was highest for sample members who attended 4-year colleges (95 percent); in contrast, contact was made with only 81 percent of sample members attending private for-profit institutions with programs of less than 2 years.

For the second BPS cohort, the contact rate for the first follow-up (BPS:96/98) was 91 percent. The overall unweighted response rate was 84 percent. Full respondents to NPSAS:96 had a contact rate almost 33 percentage points higher than NPSAS:96 nonrespondents (94 vs. 61 percent). Students from private, for-profit institutions had the lowest contact rates (79 percent for 2-year institutions and 82 percent for less-than-2-year institutions), while students from public 4-year institutions (94 percent) and private, not-for-profit 4-year institutions (94 percent) had the highest contact rates.

In the second follow-up of the second BPS cohort (BPS:96/01), the contact rate was 92 percent. The overall unweighted response rate was 88 percent. Students who had not participated in the first follow-up had a lower contact rate (81 percent) than those who had been interviewed both in NPSAS:96 and BPS:96/98 (93 percent) and those who had only been interviewed in BPS:96/98 (92 percent). Contact rates were similar across institutions, with a high of 96 percent for private not-for-profit 4-year doctorate-granting institutions and a low of 86 percent for private, for-profit less-than-2-year institutions.

The first follow-up to the third BPS cohort (BPS:04/06) reported locating 89 percent of the sample members. Of these, 81 percent were considered eligible for BPS. Among all eligible sample members (including both located and not located), the overall unweighted response rate was 80 percent; among eligible cases that were successfully located, the response rate was 90 percent. . For the second follow-up, 91 percent of the sample members were located. Eligibility did not need to be evaluated as part of BPS:04/09. The overall unweighted response rate was 82 percent; among eligible cases that were successfully located, the response rate was 90 percent.

Among those students in the first BPS cohort who were contacted for the second follow-up, the *interview rate* was 95 percent. The rate was higher for respondents to the first follow-up than for nonrespondents (96 vs. 89 percent). Interview rates were fairly similar across institutions, ranging from 91 percent for students attending private, not-for-profit less-than-2-year institutions to 96 percent for students attending private, not-for-profit 4-year institutions.

The interview rate for those contacted in the first follow-up of the second BPS cohort was 92 percent. This rate was lower for NPSAS:96 nonrespondents than for full or partial respondents (71 percent vs. 94 and 82 percent). Interview rates were much more consistent across institutions; private, for-profit 2-year institutions were the lowest at 88 percent, and private, for-profit 4-year institutions were the highest at 95 percent.

With the second follow-up to the second BPS cohort, the interview rate was 96 percent of those contacted. As with the contact rates, the interview rates varied across groups of participants. Specifically, interview rates were lower for those not interviewed in the first follow-up than for those interviewed both in the base-year study and the first follow-up and those interviewed only in the base-year study (81 percent vs. 96 and 91 percent). Interview rates varied across institutional sectors from 93 to 97 percent.

Among located eligible students, interview rates for those contacted in the first follow-up were higher for NPSAS:04 respondents (90 percent) than nonrespondents (52 percent). Similarly, BPS:04/09 interview rates were higher among first follow-up respondents (93 percent) than nonrespondents (77 percent). Across institution types, interview rates varied from 87 to 94 percent for the first follow-up and from 70 to 88 percent for the second follow-up. Of the completed interviews, 58 percent completed on the web and 42 percent were interviewer-administered (39 percent CATI and 3 percent CAPI). For the second follow-up, 64 percent of interviews were completed by web with 36 percent administered by an interviewer (32 percent CATI and 4 percent CAPI),
Table 9 summarizes the unit-level weighted response rates across three BPS administrations.

Item Nonresponse. Overall item nonresponse rates have been low across surveys (only 10 of the 363 items in BPS:96/98 and 9 of the 363 items in BPS:96/01 contained over 10 percent missing data, 7 of the more

than 400 items in BPS:04/06 and 19 of 385 items in BPS:04/09 had a total nonresponse over 5 percent). Items with the highest nonresponse rates were those for income and student loans. Many respondents were reluctant to provide information about personal and family finances or simply did not know this information.

Measurement Error. While comprehensive psychometric evaluations of BPS data have not been conducted, issues of data quality are addressed during data collection.

Cross-Interview Data Verification. During data collection, information from a prior interview (or from base-year NPSAS data) is verified or updated to ensure compatibility across survey waves. In the first follow-up of the first BPS cohort (i.e., BPS:90/92), demographic information covered in NPSAS (e.g., sex, race, and ethnicity) was verified or updated. The results indicated high reliability of these items. Prior to the full-scale second follow-up, another set of items covered in earlier rounds was verified or updated, including high school graduation status, schools attended prior to the base year, and jobs held prior to the base year. These data were also found to be reliable across survey waves. Agreement approached 100 percent on high school graduation status, 99 percent on previous attendance at postsecondary schools, and 96 percent on previous jobs.

A minimal amount of replacement was conducted on the follow-ups of the second BPS cohort. No replacement of data was conducted on the 2006 follow-up (BPS:04/06) of the third cohort because data were swapped. No replacement of data was conducted on the 2009 follow-up (BPS:04/09) because missing values were imputed from the 2006 data and data were swapped.

Table 9. Unit-level weighted response rates for BPS student surveys, by cohort: 1990–2009

Cohort	Base year Inst. level[1,2]	Base year Student level[1]	1st follow-up	2nd follow-up
1st cohort	86	84	86[3]	91
2nd cohort	91	96	80	84
3rd cohort	80	91	80[3,4]	71

[1] Base year NPSAS (analysis file) response rates.
[2] Institutional response rates for student sampling lists.
[3] Student interview response rate.
[4] Unweighted response rate.
NOTE: Follow-up response rates are overall response rates, except where noted.
SOURCE: Cominole, M., Siegel, P., Dudley, K., Roe, D., and Gilligan, T. (2006). *2004 National Postsecondary Student Aid Study (NPSAS:04) Full-Scale Methodology Report* (NCES 2006-180). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Cominole, M., Wheeless, S., Dudley, K., Franklin, J., and Wine, J. (2007). *2004/06 Beginning Postsecondary Students Longitudinal Study (BPS:04/06) Methodology Report* (NCES 2008-184). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Fitzgerald, R., Berkner, L., Horn, L.J., Choy, S.P., and Hoachlander, G. (1994). *Descriptive Summary of 1989-90 Beginning Postsecondary Students: Two Years Later: 90-92 (*NCES 94-386). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Malizio A.G. (1992). *Methodology Report for the 1990 National Postsecondary Student Aid Study (NPSAS:90) Contractor Report* (NCES 92-080). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Pratt, D.J., Whitmore, R.W., Wine, J.S., Blackwell, K.M., Forsyth, B.H., Smith, T.K., Becker, E.A., Veith, K.J., Mitchell, M., and Borman, G.D. (1996). *Beginning Postsecondary Students Longitudinal Study Second Follow-up (BPS:90/94) Final Technical Report (*NCES 96-153). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Riccobono, J.A., Whitmore R.W., Gabel T.J., Traccarella M.A., Pratt D.J., Berkner L.K., and Malizio A.G. (1997). *National Postsecondary Student Aid Study (NPSAS:96) Methodology Report* (NCES 98-073). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Wine, J.S., Heuer, R.E., Wheeless, S.C., Francis, T.L., and Dudley, K.M. (2002). *Beginning Postsecondary Students Longitudinal Study: 1996-2001 (BPS:1996/2001) Methodology Report* (NCES 2002-171). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Wine, J.S., Whitmore, R.W., Heuer, R.E., Biber, M., and Pratt, D.J. (2000). *Beginning Postsecondary Students Longitudinal Study First Follow-up 1996–98 (BPS:96/98) Methodology Report (*NCES 2000-157) National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Reinterview. All BPS field test interview activities have involved a reinterview of a subsample of respondents to the main interview to evaluate the consistency of responses to the two interviews. The interval between the initial interview and the reinterview has ranged from 3 to 14 weeks.

Across BPS data collections, each new reinterview is designed to build on previous analyses by targeting

revised items, new items, and items not previously evaluated. Reinterview analyses focus on data items that were expected to be stable for the time period between the initial interview and the reinterview. These items cover education experience; work experience (e.g., employee's primary role, future career plans, principal job's relation to education, satisfaction with principal job, and factors affecting employment goals); education finances; and living arrangements. Across cohorts and surveys, the reliability of survey items has varied in ways that are typical of the types of questions being asked and answered. Rates of agreement have tended to be high among factual questions, such as those related to enrollment history, employment, and background characteristics. Reliability has been lower among numeric responses, such as income for a calendar year and parents' income. Adjustments in question design, wording, and response options were made from field test to full-scale administration to address problems in item reliability.

When there continued to be concern for the reliability of an item, it was reevaluated in the next field test interview.

Across cohorts and surveys, the reliability of survey items has varied in ways that are typical of the types of questions being asked and answered. Rates of agreement have tended to be high among factual questions, such as those related to enrollment history, employment, and background characteristics. Reliability has been lower among numeric responses, such as income for a calendar year and parents' income. Adjustments in question design, wording, and response options were made from field test to full-scale administration to address problems in item reliability. When there continued to be concern for the reliability of an item, it was reevaluated in the next field test interview.

Item Order Effects. As needed, analyses are conducted to evaluate order effects, that is, the sequence in which questionnaire items are presented to respondents and the resulting response patterns. Discrepancies are examined and adjustments made, as needed, for the full-scale data collection. Order effects are controlled through the randomization of response options that is possible with computer-assisted interviews. Also analyzed are discrepancies of online coding procedures for postsecondary institutions, fields of study, and combined and separate industry and occupations. To achieve high data quality, expert coding personnel recode items that have been identified as inconsistent.

6. CONTACT INFORMATION

For content information about the BPS project, contact:

Matthew Soldner
Project Officer
Phone: (202) 219-7025
E-mail: matthew.soldner@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
Statistics 1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Burkheimer, Jr., G.J., Forsyth, B.H., Whitmore, R.W., Wine, J.S., Blackwell, K.M., Veith, K.J., and Borman, G.D. (1994). *Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:90/92) Final Public Technical Report* (NCES 94-369). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Cominole, M., Wheeless, S., Dudley, K., Franklin, J., and Wine, J. (2007). *2004/06 Beginning Postsecondary Students Longitudinal Study (BPS:04/06) Methodology Report* (NCES 2008-184). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Fitzgerald, R., Berkner, L., Horn, L.J., Choy, S.P., and Hoachlander, G. (1994). *Descriptive Summary of 1989-90 Beginning Postsecondary Students: Two Years Later: 90-92* (NCES 94-386). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Malizio A.G. (1992). *Methodology Report for the 1990 National Postsecondary Student Aid Study (NPSAS:90) Contractor Report* (NCES 92-080). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Pratt, D.J., Whitmore, R.W., Wine, J.S., Blackwell, K.M., Forsyth, B.H., Smith, T.K., Becker, E.A., Veith, K.J., Mitchell, M., and Borman, G.D. (1996). *Beginning Postsecondary Students Longitudinal Study Second Follow-up (BPS:90/94) Final Technical Report (*NCES 96-153). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Riccobono, J.A., Whitmore R.W., Gabel T.J., Traccarella M.A., Pratt D.J., Berkner L.K., and Malizio A.G. (1997). *National Postsecondary Student Aid Study (NPSAS:96) Methodology Report* (NCES 98-073). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Wine, J.S., Heuer, R.E., Wheeless, S.C., Francis, T.L., and Dudley, K.M. (2002). *Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Methodology Report* (NCES 2002-171). National Center for Education

Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wine, J. S,, Janson, N., and Wheeless, S.C. (forthcoming). 2004/09 Beginning Postsecondary Students Longitudinal Study (BPS:04/09) Full-scale Methodology Report. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wine, J.S., Whitmore, R.W., Heuer, R.E., Biber, M., and Pratt, D.J. (2000). *Beginning Postsecondary Students Longitudinal Study First Follow-up 1996–98 (BPS:96/98) Methodology Report* (NCES 2000-157). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Survey Design

Burkheimer, Jr., G.J., Forsyth, B.H., Wheeless, S.C., Mowbray, K.A., Boehnlein, L.M., Knight, S.M., and Veith, K.J. (1992). *Beginning Postsecondary Students Longitudinal Study Field Test Methodology Report (BPS:90/92*) (NCES 92-160). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Pratt, D.J., Wine, J.S., Heuer, R.E., Whitmore, R.W., Kelly, J.E., Doherty, J.M., Simpson, J.B., and Norman, M.C. (1998). *Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96/98) Field Test Report* (NCES Working Paper 98-11). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Wine, J., Cominole, M., Wheeless, S., Bryant, A., Gilligan, T., Dudley, K., and Franklin, J. (2006). *2004/06 Beginning Postsecondary Students Longitudinal Study (BPS:04/06) Field Test Methodology Report* (NCES Working Paper 2006-01). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Chapter 16: Baccalaureate and Beyond (B&B) Longitudinal Study

1. OVERVIEW

The Baccalaureate and Beyond (B&B) Longitudinal Study provides information concerning education and work experiences following completion of the bachelor's degree. It provides both cross-sectional profiles of the enrollment, persistence, and financial aid receipt of bachelor's degree recipients in their final year of undergraduate education and longitudinal data on their entry into and progress through graduate-level education and the workforce. Special emphasis is placed on those graduates entering public service areas, particularly teaching, and the provision of information on their entry into the job market and career path.

B&B draws the base-year data for its cohorts from the National Postsecondary Student Aid Study (NPSAS, see chapter 14). The first B&B cohort consisted of individuals who received a bachelor's degree in the 1992–93 academic year; the second cohort was formed from baccalaureate recipients in the 1999–2000 academic year; and, the third cohort consists of graduating seniors from the 2007–08 academic year (B&B:93, B&B:2000, and B&B:08, respectively). B&B expands on the efforts of the former Recent College Graduates Survey to provide unique information on educational and employment-related experiences of these degree recipients over a longer period of time. The 1992–93 cohort was followed three times over a 10-year period, in 1994, 1997, and 2003 (B&B:93/94, B&B:93/97, and B&B:93/03, respectively), so that most respondents who attended graduate or professional schools have completed (or nearly completed) their education and are established in their careers. The 1999–2000 cohort was followed only in 2001 (B&B:2000/01). The 2007–08 cohort was followed for the first time in 2009 (B&B:08/09). Eligible sample members will be interviewed again in 2012. B&B can address issues concerning delayed entry into graduate school, the progress and completion of graduate-level education, and the impact of undergraduate and graduate debt on choices related to career and family.

LONGITUDINAL SAMPLE SURVEY OF BACHELOR'S DEGREE RECIPIENTS; THREE FOLLOW-UPS OVER A 10-YEAR PERIOD:

B&B collects data from:

- Base-year NPSAS Data
- Student interviews
- Undergraduate transcripts
- Federal financial aid and loan records
- Identified newly qualified teachers

Purpose

To provide information on (1) college graduates' entry into, persistence and progress through, and completion of graduate-level education in the years following receipt of the bachelor's degree; and (2) the career paths of new teachers: retention, attrition, delayed entry, and movement within the educational system.

Components

B&B consists of base-year data collected from NPSAS: the 1992–93 NPSAS for the first B&B cohort; the 1999–2000 NPSAS for the second B&B cohort; and the 2007–08 NPSAS for the third B&B cohort (NPSAS:93, NPSAS:2000, and NPSAS:08, respectively). NPSAS data are collected in many components, including institutional records from postsecondary institutions, student interviews, and administrative federal financial aid record systems. For the first B&B cohort (consisting of 1992–93 baccalaureate recipients), the first follow-up, conducted in 1994, collected data from a student interview as well as from undergraduate

college transcripts. The second follow-up, conducted in 1997, combined a Student Interview with Department Aid Application/Loan Records data. The third follow-up, conducted in 2003, collected data on topics related to continuing education, degree attainment, employment, career choice, family formation, and finances. A second B&B cohort, consisting of 1999–2000 baccalaureate recipients, went to the field in 2000, and was followed in 2001. The first and only planned follow-up survey, it focused on time to degree completion, participation in post-baccalaureate education and employment, and the activities of newly qualified teachers. A third B&B cohort, consisting of 2007–08 baccalaureate recipients, was followed for the first time in 2009. When they are released in early 2011, the data for this follow-up will combine student interviews, undergraduate college transcript, and other administrative records. The research topics include the relationship between college graduates' coursetaking while in college and their subsequent paths into the labor market and/or through graduate school; accumulated educational debt burden of college graduates; and preparations graduates have made for elementary and secondary school teaching, particularly as compared to those of college graduates in other occupations.

Base-Year Data (from NPSAS). B&B obtains its base-year information from NPSAS. The NPSAS Student Record Abstracts (institutional records) provide major field of study; type and control of institution; attendance status; tuition and fees; admission test scores; financial aid awards; cost of attendance; student budget information and expected family contribution for aided students; grade point average; age; and date first enrolled. The base-year data also include information from NPSAS Student Interviews regarding educational level; major field of study; financial aid at other schools attended during the year; other sources of financial support; monthly expenses; reasons for selecting the school attended; current marital status; age; race/ethnicity; sex; highest degree expected; employment and income; community service; expectations for employment after graduation; expectations for graduate school; and plans to enter the teaching profession. For NPSAS:08, parental data previously collected from the Parent Interviews were captured in the Student Interview. These topics include marital status; age; highest level of education achieved; income; amount of financial support provided to the student; types of financing used to pay student educational expenses; and current employment (including occupation and industry).

B&B First Follow-up Survey. The first follow-up is conducted 1 year after the bachelor’s degree is received (e.g., 1994 for the 1992–93 cohort and 2001 for the 1999–2000 cohort). No other follow-up is being conducted of the 1999–2000 cohort. Data were collected in 2009 for the first follow-up of the 2007–08 cohort.

In the Student Interview portion of the survey, recent graduates provide information regarding employment after degree completion; job search activities; expectations for and entry into teaching; teacher certification status; job training and responsibilities; expectations/entry into graduate school; enrollment after degree; financial aid; loan repayment/status; income; family formation and responsibilities; and participation in community service. As part of the first follow-up of both the 1992–93 B&B and 2007-08 B&B cohorts, an undergraduate transcript study component collected transcripts providing information on undergraduate coursework; institutions attended; grades; credits attempted and earned; and academic honors earned. All transcript information is as reported by the institutions, converted to semester credits and a 4.0 grade scale for comparability.

B&B Second Follow-up Survey. The second follow-up of the 1992–93 B&B cohort was conducted in 1997, 4 years after the bachelor’s degree was received. Participants provided information in the Student Interview regarding their employment history; enrollment history; job search strategies at degree completion; career progress; current status in graduate school; nonfederal aid received; additional job training; entry into/persistence in/resignation from teaching career; teacher certification status; teacher career path; income; family formation and responsibilities; and participation in community service.

The second follow-up of the 1992–93 B&B cohort also included a Department Aid Application/Loan Records component to collect information on the types and amounts of federal financial aid received, total federal debt accrued, and students’ loan repayment status. One of the goals of B&B is to understand the effect that education-related debt has on graduates’ choices concerning their careers and further schooling. Data will be collected in 2012 for the second follow-up of the 2007–08 cohort.

B&B Third Follow-up Survey. The 1992–93 cohort was followed for a third time in 2003. This final interview, which was conducted 10 years following degree completion, allowed further study of the issues already addressed by the preceding follow-up studies. The 2003 interview covered topics related to continuing education, degree attainment, employment, career choice, family formation, and finances.

Additionally, respondents were asked to reflect on the value that their undergraduate education and any other education obtained since receiving the bachelor's degree added to their lives now. It also contained a separate set of questions directed at new entrants into the teacher pipeline, as well as those who were continuing in or who had left teaching since the last interview.

Periodicity

The three B&B cohorts each have their own follow-up schedule,[1] as described above. B&B cohorts alternate with Beginning Postsecondary Students (BPS) Longitudinal Study cohorts in using NPSAS surveys as their base.

2. USES OF DATA

B&B covers many topics of interest to policymakers, educators, and researchers. For example, B&B allows analysis of the participation and progress of recent degree completers in the workforce, relationship of employment to degree, income and the ability to repay debt, and willingness to enter public service-related fields. B&B also allows analysis of issues related to access into and choice of graduate education programs. Here the emphasis is on the ability, ease, and timing of entrance into graduate school, and attendance/employment patterns, progress, and completion timing once entered.

The unique features of B&B allow it to be used to address issues related to undergraduate education as well as post-baccalaureate experiences. This information has been used to investigate the relationship between undergraduate debt burden and early labor force experiences, and between undergraduate academic experiences and entry into teaching. These and other relationships can be investigated both in the short term and over longer periods.

Because B&B places special emphasis on new teachers at the elementary and secondary levels, it can be used to address many issues related to teacher preparation, entry into the profession (e.g., timing, ease of entry), persistence in or defection from teaching, and career movement within the education system.

Major issues that B&B attempts to address include:

- length of time following receipt of degree after which college graduates enter the workforce;
- type of job which graduates obtain, compared with major field of undergraduate study;
- length of time to complete degree;
- length of time to obtain a job related to respondents' field of study;
- extent to which jobs obtained relate to educational level attained by respondent;
- extent to which level of debt incurred to pay for education influences decisions concerning graduate school, employment, and family formation;
- extent to which level of debt incurred influences decisions to enter public service professions;
- rates of graduate school enrollment, retention, and completion;
- extent to which delaying graduate school enrollment influences respondent's access to and progression through advanced degree programs;
- factors influencing the decision to enroll in graduate education;
- extent to which attaining an advanced degree influences short-term and long-term earnings;
- number of graduates qualified to teach;
- extent to which degree level/profession influences rate of advancement; and
- extent to which respondents change jobs or careers.

3. KEY CONCEPTS

Some of the concepts and terms used in the B&B data collection and analysis are defined below. For more information on these and others terms used in B&B, refer to *A Descriptive Summary of 1999–2000 Bachelor's Degree Recipients, 1 Year Later, With an Analysis of Time to Degree* (Bradburn et al 2003).

[1] B&B:08 follow-up studies beyond 2009 will be conducted as funding permits.

Degree-granting Institution. Any institution offering an associate's, bachelor's, master's, doctor's, or first-professional degree. Institutions that grant only certificates or awards of any length (less than 2 years, or 2 years or more) are categorized as nondegree-granting institutions.

First Postsecondary Institution. The first institution attended by the respondent following high school and in which the respondent was enrolled for a minimum of 3 months. Institutions attended before high school graduation are included if enrollment continued after high school graduation. The first institution may or may not be the institution that granted the bachelor's degree.

Status in Teacher Pipeline. This variable measures the extent of involvement with teaching, using variables from 1994 and 1997 interviews and composites. Respondents who taught were classified as having taught (1) with certification, (2) with student teaching experience, (3) without training, or (4) with training unknown. Respondents who did not teach were classified as (1) certified, (2) having student taught, (3) having applied for teaching jobs, (4) having considered teaching, or (5) having no interest in or taken no action toward teaching. An additional category of respondents who had become certified but whose teaching status was unknown was identified. All of these categories are combined in various ways throughout reports, depending on the context of the particular analysis.

Dependency Level. If a student is considered financially dependent, the parents' assets and income are considered in determining aid eligibility. If the student is financially independent, only the student's assets are considered, regardless of the relationship between student and parent. The specific definition of dependency status has varied across surveys. In the 1999-2000 NPSAS, a student is considered independent if (1) the institution reports that the student is independent or (2) the student meets one of the following criteria: (a) is age 24 or older as of 12/31/1999; (b) is a veteran of the U.S. Armed Forces; (c) is an orphan or ward of the court; (d) is enrolled in a graduate or professional program beyond a bachelor's degree; (e) is married; or (f) has legal dependents other than a spouse.

4. SURVEY DESIGN

Target Population

All postsecondary students in the 50 states, the District of Columbia, and Puerto Rico who completed a bachelor's degree in the 1992–93 academic year, spanning July 1, 1992, to June 30, 1993 (first B&B cohort); in the 1999–2000 academic year, spanning July 1, 1999, to June 30, 2000 (second B&B cohort) or in the 2007–08 academic year, spanning July 1, 2007, to June 30, 2008 (third B&B cohort). Students from U.S. Service Academies are excluded because they are not part of NPSAS, from which B&B draws its samples.

Sample Design

Members of the B&B cohort are identified during the NPSAS year that serves as the base year for the longitudinal study: NPSAS:93 for the first B&B cohort, NPSAS:2000 for the second B&B cohort, and NPSAS:08 for the third B&B cohort. (See chapter 14 for a description of the NPSAS sample design.) The B&B cohorts consist of students who have completed the NPSAS interview and have been identified as baccalaureate recipients. The B&B:93 and B&B:08 cohorts also consist of those NPSAS:93 and NPSAS:08 nonrespondents, respectively, who are potentially eligible for B&B and for whom there are at least some data (either from institutional records or computer-assisted telephone interviewing [CATI]). The NPSAS sampling design is a two-stage design in which eligible institutions are selected first and then eligible students are selected from the eligible participating institutions.

Selection of Institutions

The institution-level sampling frames for NPSAS:93, NPSAS:2000, and NPSAS:08 were constructed from the 1990–91 Integrated Postsecondary Education Data System (IPEDS) file, the 1998–99 IPEDS file, and the 2005–06 IPEDS file, respectively. The resulting sampling frames contained 10,140 potentially eligible institutions for NPSAS:93, 6,420 institutions for NPSAS:2000, and 6,780 institutions for NPSAS:08.

Geographic areas defined by three-digit postal zip codes were used as the basis for creating primary sampling units (PSUs) of nearly equal sizes to ensure statistical efficiency (the three-digit code comes from the first three digits of a zip code, and designates either a sectional center facility or a main post office). All institutions within the sample PSUs were then combined into a single frame, stratified by 22 strata. The variables used to define the strata were institutional control, highest level of offering, and the percentage of baccalaureate degrees awarded in education.

For the NPSAS:93 sample, a sample of 1,360 institutions (720 from the certainty PSUs and 640 from the noncertainty PSUs) was selected for the primary sample from the IPEDS frame. For the NPSAS:2000

sample, a sample of 1,080 institutions (290 from the certainty PSUs and 800 from the noncertainty PSUs) was selected for the primary sample from the IPEDS frame. For the NPSAS:08 sample, the final sample included 1,960 institutions, and of those, about 1,960 were selected to participate in NPSAS:08.[2]

Selection of Students

Base-Year Survey. To create the NPSAS student sampling frame, each sample institution was asked to provide a list of all students enrolled during the NPSAS year (July 1, 1992 to June 30, 1993 for the first B&B cohort; July 1, 1999 to June 30, 2000 for the second B&B cohort; and July 1, 2007 to June 30, 2008 for the third B&B cohort) and those eligible to receive a baccalaureate degree at some point during that year, according to criteria provided to the institutions. Stratified systematic sampling was used to facilitate sampling from lists. For each sample institution, student sampling rates were determined for each of five student sampling strata:

- business major baccalaureates;
- other baccalaureate recipients;
- other undergraduates, including enrollees at less-than-4-year institutions;
- graduate students; and
- first-professional students.

The sampling rates depended on the overall population sampling rates for the five types of students, the probability of selecting the institution, and a requirement for a minimum of 40 sample students per institution whenever possible. Sample institutions identified those students eligible to receive the bachelor’s degree during the academic year for inclusion in each B&B cohort. In addition, those students who indicated in the CATI that they had received a baccalaureate degree during the 1992–93 academic year were also included in the B&B:93 cohort. From the NPSAS:93 sample, 16,320 baccalaureate degree recipients were identified for participation in B&B:93. From the NPSAS:2000 sample, 16,620 baccalaureate degree recipients were identified for participation in B&B:2000.[3]

[2] Additional details on the NPSAS:08 sample will become available upon release of the relevant study data.
[3] The final number of students in the B&B:08 sampling frame will be determined upon release of the relevant NPSAS:08 study data.

First Follow-up Survey.[4] About 16,320 baccalaureate degree recipients were identified for inclusion in the B&B:93 cohort from institutionally-provided lists of students who were eligible for graduation or who indicated in the CATI interview that they had graduated in the 1992–93 academic year. All 11,810 of the identified students who completed the NPSAS:93 interview were retained for the B&B:93/94 sample. Also retained were 370 student nonrespondents for whom NPSAS parent data were available that indicated that the student received the bachelor’s degree during 1992–93. Additionally, a 10 percent subsample of the remaining eligible cases with at least some data was included, for a total of 12,730 eligible cases. It became apparent during data collection that many of the nonrespondents and potentially eligible cases were actually ineligible. Because of the costs associated with the ineligible students, only a subsample of the nonrespondents and potentially eligible students was selected, reducing the sample size to 12,480 in B&B:93/94.

The respondent universe for the B&B:2000/01 follow-up survey consisted of all students who attended postsecondary educational institutions between July 1, 1999, and June 30, 2000, in the United States and Puerto Rico, and who received or expected to receive bachelor’s degrees during this time frame. Approximately 11,700 confirmed and potentially eligible bachelor’s degree recipients were selected for participation in B&B:2000/01. Of these, about 70 were determined during the follow-up survey to be ineligible. From the remaining nearly 11,630 eligible sample members, about 10,030 were located and interviewed in the follow-up survey.

Second Follow-up Survey. B&B:93/94 included a transcript component, which was used to determine eligibility of the base-year nonrespondents for the B&B:93/97 follow up. After data collection was complete for the first follow-up, additional ineligible cases were found in the cohort based on information obtained from the transcript data. Sample members were retained for follow-up in later rounds if they were found to be eligible in either the CATI or the transcript component. In total, 11,190 cases were retained for the second follow-up (B&B:93/97). Specifically, B&B:93/97 included 10,080 CATI-eligible cases, 1,090 transcript-eligible cases, and 20 cases for which eligibility was unknown for both components.

Third Follow-up Survey. All 10,090 B&B:93/97 respondents were included in the B&B:93/03 sample.

[4] The discussion of the follow-up surveys pertains to the first two B&B cohorts only; the first follow-up of B&B:08 took place in 2009.

However, because it is more difficult and expensive to locate and interview prior nonrespondents, a subsample of only about one-third, or 360, of B&B:93/97 nonrespondents was included. After removing 10 cases identified as deceased, the final sample for B&B:93/03 was 10,440.

Data Collection and Processing

B&B surveyed its first cohort—1992–93 bachelor's degree recipients—in 1994, approximately 1 year after graduation, and again in 1997. Both of these follow-up surveys were administered by the National Opinion Research Center (NORC) at the University of Chicago. The third follow-up was conducted in 2003 by Research Triangle Institute (RTI).

The first follow-up of the 1999–2000 cohort (B&B:2000/01) was conducted in 2001 by RTI. This cohort of students was first interviewed in NPSAS:2000, the base-year study for this cohort. B&B:2000/01 is the only planned follow-up of this cohort.

The first follow-up of the third cohort (B&B:2008/09) was conducted in 2009 by RTI. This cohort of students was first interviewed in NPSAS:08, the base-year study for this cohort.

Reference dates. In the first follow-up of the 1992–93 cohort, respondents were asked to provide their current enrollment status, employment status, and marital status as of April 1994. Similarly, respondents to the second and third follow-ups reported their status as of April 1997 and April 2003. For the follow-up of the 2000–01 cohort, respondents were asked to provide their current enrollment status, employment status, and marital status as of April 2001.

Data collection. Data are collected through student interviews and college transcripts. The data collection procedures for the follow-ups of the first and second B&B cohorts are described below.

Student interview. The first follow-up student interview (B&B:93/94) was administered between June and December 1994. Sample members were initially mailed a letter containing information about the survey and a toll-free number they could call to schedule interviews. CATI began approximately 1 week later and was conducted in two waves. Wave 1 consisted of students who were respondents in the 1992–93 NPSAS or for whom parent data were available. Wave 2 consisted of students who were nonrespondents in the 1992–93 NPSAS and for whom no parent data were available. NPSAS respondents who were identified as potentially eligible for B&B during the NPSAS data processing phase were also included in Wave 2.

Telephone interviewing continued for a period of 16 weeks. All cases still pending after this time were sent to field interviewers to gather in-person information. A 14-call maximum was set, with a call defined as contact with the sample member, another person in the sample member's household, or an answering machine. After 14 calls, attempts to contact the sample member by telephone were terminated and the case was sent to field interviewers.

Methods of refusal conversion were tailored to address the reasons each member had given for nonparticipation, as determined by reviewing the call notes. Letters were sent to sample members addressing the specific reasons for their refusal (too busy, not interested, confidentiality issues, etc.). Following these mailings, a final phone interview was attempted from the central CATI site. Continuing refusals were forwarded to the field to be contacted in person by a field interviewer. The field staff was successful in completing 3,050 (82 percent) of these cases.

The data collection procedures for the first (and only) follow-up of the second B&B cohort were similar to those for the first cohort, consisting almost exclusively of CATI interviews, and concluding with refusal conversion procedures to gain cooperation from telephone nonrespondents.

The second follow-up student interview (B&B:93/97) was administered between April and December 1997. Sample members were initially mailed a letter and informational leaflet containing information about the survey and a toll-free number and/or e-mail address through which they could obtain further information, schedule an interview, or provide an updated phone number. CATI began approximately 1 week later and continued for 16 weeks. Cases pending at the end of this time were sent to field interviewers and worked from July through December 1997. Phone interviewers made 13, rather than 14, attempts to contact sample members. If phone interviewers were not successful after 13 attempts, the case was forwarded to telephone case management specialists before being sent to field interviewers.

Slight modifications were also made to the methods used to locate sample members. Prior to the beginning of CATI, all cases had been sent to a credit bureau database service to obtain updated phone and address information about each sample member. Telephone numbers were also available from the previous interview (B&B:93/94 or NPSAS:93) and the National

Change of Address (NCOA)/Telematch update service that NORC had used for all main survey respondent data in February 1996, prior to the start of the field test. The "best" phone number was assumed to be the number most recently obtained.

Additional information used by locating specialists (in order of use) was as follows: (1) all respondent-generated information (e-mails, address corrections from the U.S. Post Office, any previously acquired respondent phone numbers); (2) the last known telephone number of the parent(s); (3) graduate schools (if applicable); (4) undergraduate institutions/alumni associations; (5) the other two credit bureau updating services; (6) a military locating service, if applicable; and (7) the Department of Motor Vehicles in the state that issued the respondent's last known driver's license.

A total of 1,680 respondents (15 percent of the total eligible sample) refused to complete the B&B:93/97 interview at some point in the process. After a 2-week "cooling off" period, these cases were contacted by trained interviewers experienced in refusal conversion. The CATI refusal converters were able to complete 340 of the refusal cases. Continuing refusals were forwarded to the field to be contacted in person by a field interviewer. A total of 3,990 cases (36 percent of the total sample) were sent to the field staff, which was successful in completing 2,950 (74 percent) of these cases.

The third follow-up interview (B&B:93/03) started in February 2003. For the first time, respondents were offered the opportunity to conduct their own interview via the Internet. A single, web-based interview was designed and programmed for use as a self-administered interview, a telephone interview, and an in-person interview. In addition, a website was developed to launch the self-administered interview, to provide additional study information, and to collect updated student locating information.

Three weeks after the self-administered interview was made available to sample members in February 2003, telephone interviewing began with those sample members who had not yet completed the self-administered interview. About 3 months after the start of telephone interviewing, field interviewers began tracing and interviewing nonrespondents whose last known address was in one of 30 geographic clusters. From the starting sample of 10,440, about 40 individuals were found to be deceased and another 10 were determined to be ineligible. The unweighted locating rate among the remaining sample members was 93 percent. Of those located, 92 percent completed the interview, for an overall unweighted response rate of 86 percent. Among respondents, 38 percent completed the self-administered interview on the Internet, 57 percent completed a telephone interview, and the remaining 5 percent were interviewed in person.

Incentives were offered to sample members at two different points during data collection. First, sample members were offered a $20 cash incentive for completing the self-administered interview within the first 3 weeks of data collection, prior to the start of telephone interviewing. Of those who completed the self-administered interview, 47 percent did so during the incentive period. Additionally, an incentive was used to reduce nonresponse among four groups: those who refused to be interviewed, those who could not be reached by telephone, those for whom only a contact person could be reached, and those who started but did not finish the self-administered interview. Overall, 55 percent of sample members falling into one of the four groups completed the interview following the offer of a nonresponse incentive.

Among the telephone interviewers was a group of refusal conversion specialists trained in converting sample members who have refused to complete the interview. From the point when a sample member refused, the case was handled only by these conversion specialists. In B&B:93/03, slightly less than 10 percent of sample members ever refused to participate in the interview. Of these sample members, 49 percent eventually completed the interview.

Transcript component. In addition to data gathered from sample members, the first B&B follow-up included a transcript component that attempted to capture student-level coursetaking and grades for eligible sample members. Transcripts were requested for all sample members from the NPSAS schools that awarded them their bachelor's degrees.

Data collection for the transcript component began in August 1994, when request packets were mailed to all 720 NPSAS sample schools from which B&B sample members had graduated. In addition to student transcripts, schools were asked to provide a course catalog and information on their grading and credit-granting systems and their school term. A transcript was requested for all 12,480 students in the B&B sample, although not all transcripts were coded due to sample member ineligibility. Prompting of nonresponding schools began in September 1994 by the telephone center, and attempts were made to address any concerns of school staff regarding confidentiality or the release of transcripts.

The design of the transcript processing system capitalized on work done in previous NORC studies. The process and flow system, however, was changed in four significant areas. First, since the sample of schools from which transcripts were collected was known, the system was designed around the school as the primary unit rather than around the student. Second, transcripts were entered after all school-level information about schedule, grading, and credit-granting systems had been collected and verified. The system enforced these parameters and ensured that the transcripts were internally consistent within the school. Third, the transcript coders worked with the full transcript when entering and coding courses. This allowed the coders to view each entry in context and make intelligent, informed decisions when they encountered difficult situations. Finally, the system was designed so that course-level information within schools was entered only once; subsequent duplicate course entries were selected by the coder from a dynamic school-level list of all courses entered from previous transcripts. If a course failed to match a preexisting entry, the coder searched the school-level table to see if other courses existed for the abbreviation. If a course was not in the table, the coder entered the full course title, the number of credits, and the grade.

Editing. Various edit checks, including CATI edits, have been used in processing B&B data; however, they have not been documented in B&B methodology reports for the base year and first two follow-ups of the B&B:93 cohort.

The coding and editing procedures for the B&B:93 cohort's third follow-up (B&B:93/03) fell into two categories: (1) online coding and editing performed during data collection and (2) post-data collection editing. All data collection for B&B:93/03 used one major system—a web instrument—that included edit checks to ensure that the data collected were within valid ranges. To the extent feasible, this system incorporated across-item consistency edits. Whereas more extensive consistency checks would have been technically possible, the use of such edits was limited to prevent excessive respondent burden.

Both during and after data collection, edit checks were performed on the B&B:93/03 data file to confirm that the intended skip patterns were implemented during the interview. Special codes were added after data collection, as needed, to indicate the reason for missing data. In addition, skip-pattern relationships in the database were examined by methodically running cross-tabulations between gate items and their associated nested items.

For the B&B:2000/01 data, the coding and editing procedures fell under the same two categories as above. During data collection, online coding and editing were performed, requiring CATI range and consistency checks. After data collection, edit checks were performed to verify that the database reflected appropriate skip patterns.

Estimation Methods

Weighting is used in B&B to adjust for sampling and unit nonresponse. Imputation is used to estimate baseline weights from NPSAS when these data are missing and to estimate values when the value is missing; however, no imputation was performed on data collected in the first and second follow-ups of B&B:93. Weighting procedures for the first and second cohorts are described below.

Weighting. For the first B&B cohort's first follow-up, the final weights were calculated by modifying baseline weights in NPSAS:93 to adjust for nonresponse in the B&B:93/94 survey and for tighter eligibility criteria in the B&B sample. NPSAS:93 sample development and weights calculation documentation can be found in the *Sampling Design and Weighting Report for the 1993 National Postsecondary Student Aid Study* (Whitmore, Traccarella, and Iannacchione 1995).

After verifying sample eligibility against transcript data, B&B sample members were stratified according to institutional type and student type. These strata reflected the categories used in NPSAS:93, with some modifications. NPSAS:93 categorized schools into 22 institutional strata based on highest degree offered, control (public or private), for-profit status, and the number of degrees the institution awarded in the field of education (with schools subsequently designated "high ed" or "low ed"). For weighting purposes, these 22 institutional strata were collapsed in B&B into the 16 that granted baccalaureate degrees. The six NPSAS strata representing 2-year or less-than-2-year institutions were reclassified in B&B according to control and included in the correlative "4-year, bachelor's, low ed" stratum. This affected a total of 19 cases. The five student types originally identified in the NPSAS were collapsed in B&B into three types: baccalaureate business majors, baccalaureate other majors, and baccalaureate field unknown, resulting in 48 total cells.

Baseline weights for all B&B-eligible students were adjusted for final degree totals. Control totals for baccalaureate degrees awarded were calculated based on the IPEDS Completions file for academic year 1992–93. The NPSAS institution sample frame was matched to the IPEDS file, and the total number of

baccalaureate degrees awarded was calculated by institutional stratum. An adjusted weight was calculated for each case by multiplying the NPSAS base weight by the ratio of the sum of degrees awarded to the sum of the base weights for the appropriate institutional stratum. This weight became the B&B base weight.

In order to make nonresponse adjustments for weights, adjustment cells were created by cross-classifying cases by institutional stratum and student type. Each cell was checked to verify that it met two conditions: (1) the cell contained at least 15 students; and (2) the weighted response rate for the cell was at least two-thirds (67 percent) of the overall weighted response rate. Any cells that did not meet both conditions were combined into larger cells by combining two student-type cells (baccalaureate business majors and "all other degrees") within the same institutional stratum. If this larger cell still did not meet the criteria specified above, all three student types from that institutional stratum were combined. Once all cells were defined, the B&B base weight variable (derived above) was multiplied by the inverse of the weighted response rate for the cell.

Final weights for the second follow-up (B&B:93/97) were calculated using a two-step process. First, the base weight calculated for the B&B:93/94 sample was adjusted for non-response to the B&B:93/97 survey. Next, the panel weight was calculated for respondents who participated in all three of the B&B surveys (NPSAS:93, B&B:93/94, and B&B:93/97). The 16 institutional-type and 3 student-type strata were used again, with the same process described previously.

The base weights for the third follow-up (B&B:93/03) were calculated adjusting for the subsample of nonresponding students from B&B:93/97 that were included in the B&B:93/03 survey. The cross-sectional weights for the third follow-up were developed by analyzing 8,970 respondents to the B&B:93/03 interview, using three steps of nonresponse adjustment: inability to locate the student, refusal to be interviewed, and other noninterview adjustments. All nonresponse adjustments were fitted using RTI's proprietary generalized exponential modeling (GEM) procedure. To detect important interactions for the logistic models, a Chi-squared automatic interaction detection (CHAID) analysis was performed on the predictor variables. In addition, a longitudinal weight was constructed for analyzing students who participated in all four interviews—NPSAS:93, B&B:93/94, B&B:93/97, and B&B:93/03. This weight was constructed by applying an additional nonresponse adjustment to the final B&B:93/03 cross-sectional weight. As for the other models, CHAID was used to determine the interaction segments, and GEM was used to determine the adjustment factor.

For the second B&B cohort's first follow-up (B&B:2000/01), weights were obtained in the following manner: the sample design included the first two stages of NPSAS:2000 sample design and an additional B&B:2000/01-specific stage in which a subsample was selected from confirmed and potential baccalaureate recipients identified at the end of the NPSAS:2000 sample. All confirmed baccalaureate recipients were selected into the B&B:2000/01 sample, while nonresponding potential baccalaureate recipients were randomly selected according to probabilities based on a measure of size, which was the estimate of the NPSAS:2000 study weight at the time of sample selection. Once the B&B:2000/01 sample had been selected, initial weights were obtained by adjusting the NPSAS:2000 study weights for both the B&B subsample design and the presence of study-ineligible individuals in the B&B sampling frame. Similar to the first cohort, obtaining the final weights involved using CHAID to determine the interaction segments and GEM to determine the adjustment factor.

For the third B&B cohort's first follow-up (B&B:08/09), weights were obtained in the following manner: the sample design included the first two stages of NPSAS:08 sample design and an additional B&B:08/09-specific stage in which a subsample was selected from confirmed and potential baccalaureate recipients identified at the end of the NPSAS:08 sample and the B&B transcript collection. All confirmed baccalaureate recipients were selected into the B&B:08/09 sample, while nonresponding potential baccalaureate recipients were randomly selected according to probabilities based on a measure of size, which was the estimate of the NPSAS:08 study weight at the time of sample selection. Once the B&B:08/09 sample had been selected, initial weights were obtained by adjusting the NPSAS:08 study weights for both the B&B subsample design and the presence of study-ineligible individuals in the B&B sampling frame. Obtaining the final weights involved using CHAID to determine the interaction segments and GEM to determine the nonresponse and calibration (poststratification) adjustment factors.

Imputation. The sample for the first B&B cohort (B&B:93) included 23 eligible cases in which the baseline weight from the 1992–93 NPSAS was equal to zero. Weights for these cases were imputed using the average of all nonzero baseline weights within the same institution at which the baccalaureate degree was attained. One of the cases with a missing weight

happened to be the only representative of that institution. The baseline weight was imputed for this case by using the average across all nonzero weights within the same institutional stratum and student type cell.

There was no other imputation of data items in the base-year and first two follow-ups of B&B:93.

In the third follow-up (B&B:93/03), key variables to be used in cross-sectional estimates were imputed. The imputations were performed in three steps. In the first step, the interview variables were imputed. Then, using the interview variables, including the newly imputed variable values, the set of derived variables was constructed. In the final step, the derived variables were imputed again. Only one continuous variable was imputed. Income from work in 2002 had a weighted mean of $50,846 ($n$ = 8,540) prior to imputation and a weighted mean of $50,961 ($n$ = 8,810) after imputation.

Weighted sequential hot deck imputation was selected for B&B:93/03 in part because it has the advantage of controlling the number of times a respondent record can be used for imputation and gives each respondent record the chance to be selected for use as a hot deck donor. To implement the procedure, imputation classes and sorting variables relevant to each item being imputed were defined. If more than one sorting variable was used, a serpentine sort was performed in which the direction of the sort (ascending or descending) changed each time the value of the previous sorting variable changed. The serpentine sort minimized the change in student characteristics every time one of the sorting variables changed its value.

Imputation classes for the B&B:93/03 interview variables, and some of the derived variables, were developed using a CHAID analysis where only respondent data were modeled. The CHAID segmentation process first divided the data into groups based on categories of the most significant predictor of the item being imputed, and then split each of the groups into smaller subgroups based on the other predictor variables. The CHAID process also merged categories for variables found not to be significantly different. This splitting and merging process continued until no additional statistically significant predictors were found. Imputation classes for B&B:93/03 were then defined from the final CHAID segments.

No imputations were performed for the second B&B cohort.

Imputations will be done for the third B&B cohort for the interview variables. The imputed values will then be used to form derived variables. Similar to B&B:93/03, weighted sequential hot deck will be used with imputation classes and serpentine sorting. SAS Enterprise Miner will be used to form the imputation classes using a tree algorithm similar to CHAID.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Taylor Series approximations and Balanced Repeated Replication (BRR) are used to estimate standard errors in the first and second cohorts of B&B and Taylor series approximations and bootstrap replication will be used for the third cohort.

Nonsampling Error

The majority of nonsampling errors in B&B can be attributed to nonresponse. Other sources of nonsampling error include the use of ambiguous definitions; differences in interpreting questions; an inability or unwillingness to give correct information; mistakes in recording or coding data; and other instances of human error that occur during the multiple stages of a survey cycle. Different types of nonsampling errors are described below.

Coverage error. The B&B sample is drawn from NPSAS. Consequently, any coverage error in the NPSAS sample will be reflected in B&B. (Refer to chapter 14 for coverage issues in NPSAS.)

Nonresponse error. Overall response rates were generally high for the follow-up surveys. Unit and item nonresponse data are broken down below.

Unit nonresponse. Of the 12,480 cases originally included in the first B&B sample, 1,520 were determined during the interview process to be ineligible or out of scope (primarily because their date of graduation fell outside the July 1–June 30 window). A total of 10,960 cases were considered to be eligible during the interviewing period of the first B&B follow-up, and interviews were completed with 10,080 of these respondents, representing a 92 percent unweighted response rate.

Response rates were even higher for transcript collection. In all, 630 of 640 eligible schools complied with the request for transcripts, providing transcripts for 10,970 of the 12,480 cases—a 98 percent response rate.

In the second follow-up, of the 11,190 cases identified as eligible B&B sample members, 30 were subsequently found to be out of scope or ineligible (29 were sample members who had died since 1993, and one case was identified as ineligible when it was determined the respondent had never received a baccalaureate degree). Interviews were completed with 10,970 of the 11,220 in-scope cases, for a final unweighted response rate of 90 percent. While response rates were similar across many demographic subgroups, some distinctive differences exist. Response rates decreased slightly with age (93 percent of those under 26 compared to 91 percent of those over 30 participated), but participation among males and females was approximately equal. Response rates were also similar among Whites, Blacks, and American Indians (ranging from 90 percent to 92 percent), but substantially lower for Asians/Pacific Islanders (only 82 percent) and those identifying themselves as "other" (74 percent).

In the third follow-up, about 40 individuals from the starting sample of 10,440 were found to be deceased and another 10 were determined to be ineligible. Of the B&B:93/03 sample members who were eligible to participate, 8,970 were interviewed, for an overall unweighted response rate of 87 percent (83 percent weighted). The rate of population coverage varies by type of institution: the rate is higher for public institutions than for private institutions, and higher for institutions offering a master's or doctoral degree than for those offering a bachelor's or less or a first-professional degree.

In the second B&B cohort's follow-up (B&B:2000/01), about 760 individuals from the starting sample of about 11,700 were not located, about 190 were considered "exclusions," and about 70 were deemed ineligible. A total of about 10,030 (of the approximately 11,520 remaining cases after removing the exclusions) were interviewed. An unweighted CATI response rate for B&B:2000/01 was 86 percent. The weighted overall CATI response rate was 75 percent.

Table 10 summarizes the unit-level (respondent-level) and overall-level (school-level) weighted response rates across B&B administrations.

Item nonresponse. Of the more than 1,000 variables included in the final dataset for the first cohort, 68 contain a response rate of less than 90 percent. The highest nonresponse rate was for items involving recollection of test scores and dates. Respondents also had difficulty recalling detailed information about undergraduate loans and loan payments when they had more than three loans. The two primary sections of the survey, concerning postbaccalaureate education and employment, had very low rates of nonresponse.
For the second cohort, efforts were made to encourage responses to all interview questions and to limit indeterminates, defined as a "don't know" response or a refusal to answer a question. As a result of these efforts, item nonresponse throughout the interview was low, with only 6 of 556 items having indeterminate response rates above 10 percent.

Measurement error. Three sources of measurement error identified in B&B are respondent error, interviewer error, and error in the coding of course data from transfer schools where no school-level data were available.

Respondent error. Several weeks after the first follow-up interview of the 1992–93 cohort (B&B:93/94), a group of 100 respondents was contacted again for a reinterview. These respondents were asked a subset of the items included in the initial interview to help assess the quality of these data. The results indicate that the questions elicited similar information in both interviews. Ninety-two percent of respondents gave consistent responses when asked if they had taken any courses for credit since graduating from college. Among the 8 percent with inconsistent responses, most had a short enrollment spell that they mentioned in the initial interview but not in the reinterview.

Ninety-six percent of respondents gave consistent information in both interviews when asked whether they had worked since graduation. Almost three-quarters of respondents gave the same number in both interviews when asked about the number of jobs they held since graduation; 26 percent gave inconsistent responses. Upon scrutiny, many of these discrepancies resulted from jobs held around the time of graduation that were reported in just one of the interviews. Although respondents were asked to include jobs that began before graduation if they ended after graduation, confusion over whether to include such jobs accounted for many of the inconsistencies noted in the reinterview. The 1993–94 B&B field test also included a reinterview study (see *Measurement Error Studies at the National Center for Education Statistics* [Salvucci et al. 1997]).

Interviewer error. The monitoring procedure for statistical quality control used in B&B extends the traditional monitoring criteria (which focus specifically on interviewer performance) to an evaluation of the data collection process in its entirety. This improved monitoring system randomly selects active work stations and segments of time to be monitored, determines what behaviors will be monitored and

precisely how they will be coded, and allows for real-time performance audits, thereby improving the timeliness and applicability of corrective feedback and enhancing data quality. Results for the first follow-up of the 1992–93 B&B cohort revealed a low rate of interviewer error, about three errors for every 100 minutes monitored.

Quality control procedures are also established for field interviewing. The first two interviewer-administered completed questionnaires are sent to a field manager for editing. These cases are edited, logged, and reported weekly, and appropriate feedback is given to the interviewer. Additionally, 10 percent of these cases, whether administered over the phone or in person, are validated by field managers. When deemed necessary, the field managers continue to edit additional cases to monitor data quality. The need for additional monitoring is based on the field manager's subjective judgment of the field interviewer's skill level. As with the edited cases, validated cases are logged and reported weekly.

Transfer school course coding. The first follow-up of the 1992–93 B&B cohort included a transcript data collection. Although transcripts were requested only from the institution awarding the baccalaureate degree, transcripts from previous transfer schools were often attached. Course data from these transfer school transcripts were coded, but no attempt was made to collect additional information from these schools. Due to the lack of school-level information on the 1,938 transfer schools involved, data from these transcripts are not of the same quality as data coded from the baccalaureate institution's transcripts.

Table 10. Unit-level and overall-level weighted response rates for selected B&B surveys, by data collection wave and cohort

	Unit-level weighted response rate				
	Base year Inst. level[1]	Base year Student level	1st follow-up	2nd follow-up	3rd follow-up
1992–93 student cohort	88	76	92[2]	90[2]	83
1999–2000 student cohort	91[3]	72[3]	82	†	†
2007-08 student cohort	90	64	†	†	†
	Overall-level weighted response rate				
	Base year Inst. level[1]	Base year Student level	1st follow-up	2nd follow-up	3rd follow-up
1992–93 student cohort	88	67	81[2]	80[2]	74
1999–2000 student cohort	91[3]	66[3]	74	†	†
2007-08 student cohort	90	86[4]	†	†	†

†Not applicable.
[1] Base year institutional response rates for student sampling lists.
[2] Unweighted response rate.
[3] NPSAS:2000 response rate (includes less-than 4-year institutions).
[4] Response rates calculated for study respondents as defined for NPSAS:08
NOTE: Follow-up response rates are for student interviews.
SOURCE: Loft, J.D., Riccobono, J.A., Whitmore, R.W., Fitzgerald, R.A., and Berkner, L.K. (1995). *Methodology Report for the National Postsecondary Student Aid Study, 1992–93* (NCES 95-211). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Green, P.J., Meyers, S. L., Giese, P., Law, J., Speizer, H.M., and Tardino, V.S. (1996). *Baccalaureate and Beyond Longitudinal Study: 1993/94 First Follow-up Methodology Report* (NCES 96-149). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Green, P., Meyers, S., Veldman, C., and Pedlow, S. (1999). *Baccalaureate and Beyond Longitudinal Study: 1993/97 Second Follow-up Methodology Report* (NCES 1999-159). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Riccobono, J.A., Cominole, M.B., Siegel, P.H., Gabel, T.J., Link, M.W., and Berkner, K.L. (2002). *National Postsecondary Student Aid Study 1999–2000 (NPSAS:2000) Methodology Report* (NCES 2002-152). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Charleston, S., Riccobono, J., Mosquin, P., and Link, M. (2003). *Baccalaureate and Beyond Longitudinal Study: 2000–01 (B&B:2000/01) Methodology Report* (NCES 2003-156). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC. Wine, J.S., Cominole, M.B., Wheeless, S., Dudley, K., and Franklin, J. (2005). *1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03)* (NCES 2006-166). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

6. CONTACT INFORMATION

For content information on B&B, contact:

Ted Socha
Phone: (202) 502-7383
E-mail: ted.socha@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Bradburn, E.M., Berger, R., Li, X., Peter, K., and Rooney, K (2003). *A Descriptive Summary of 1999–2000 Bachelor's Degree Recipients, 1 Year Later, With an Analysis of Time to Degree* (NCES 2003-165). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Charleston, S., Riccobono, J., Mosquin, P., and Link, M. (2003). *Baccalaureate and Beyond Longitudinal Study: 2000-01 (B&B:2000/01) Methodology Report* (NCES 2003-156). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Green, P., Meyers, S., Veldman, C., and Pedlow, S. (1999). *Baccalaureate and Beyond Longitudinal Study: 1993/97 Second Follow-up Methodology Report* (NCES 1999-159). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Loft, J.D., Riccobono, J.A., Whitmore, R.W., Fitzgerald, R.A., and Berkner, L.K. (1995). *Methodology Report for the National Postsecondary Student Aid Study, 1992–93* (NCES 95-211). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, National Center for Education Statistics (2001). *National Postsecondary Student Aid Study: Student Financial Aid Estimates for 1999–2000* (NCES 2001-209). Washington, DC:U.S. Government Printing Office.

Riccobono, J.A., Cominole, M.B., Siegel, P.H., Gabel, T.J., Link, M.W., and Berkner, K.L. (2002). *National Postsecondary Student Aid Study 1999-2000 (NPSAS:2000) Methodology Report* (NCES 2002-152). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wei, C.C., Berkner, L., He, S., Lew, S., Cominole, M., and Siegel, P. (2009). *2007–08 National Postsecondary Student Aid Study (NPSAS:08): Student Financial Aid Estimates for 2007–08: First Look* (NCES 2009-166). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wine, J.S., Cominole, M.B., Wheeless, S., Dudley, K., and Franklin, J. (2005). *1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03)* (NCES 2006-166). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Survey Design

Green, P.J., Speiger, H.M., and Campbell, B.K. (1994). *Baccalaureate and Beyond First Follow-up Field Test Report, 1993 (B&B:93/94)* (NCES 94-371). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Veldman, C., Green, P.J., Meyers, S., Chuchro, L., and Giese, P. (1997). *Baccalaureate and Beyond Longitudinal Study: Second Follow-up Field Test Report, 1996* (NCES 97-261). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Whitmore, R.W., Traccarella, M.A., and Iannacchione, V.G. (1995). *Sampling Design and Weighting Report for the 1993 National Postsecondary Student Aid Study*. Research Triangle Park, NC: Research Triangle Institute.

Data Quality and Comparability

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Chapter 17: Survey of Earned Doctorates (SED)

1. OVERVIEW

The Survey of Earned Doctorates (SED) is an annual census of new doctorate recipients from accredited colleges and universities in the United States. The SED is conducted by the National Opinion Research Center (NORC) at the University of Chicago and is funded by six federal agencies: the National Science Foundation (NSF), the lead sponsor; the Department of Education; the Department of Agriculture (USDA); the National Institutes of Health (NIH); the National Endowment for the Humanities; and the National Aeronautics and Space Administration.

Only research doctorates—primarily the Ph.D., Ed.D., and D.Sc.—are counted in the SED. Professional doctorates (e.g., M.D., J.D., Psy.D.) are excluded. While graduate schools are responsible for distributing the survey forms to students, the surveys are completed by the doctorate recipients themselves. The surveys collect information on recipients' demographic characteristics, educational history (from high school to doctorate), sources of graduate school support, debt level, and postgraduation plans.

The first SED was conducted during the 1957–58 academic year. In addition to housing the results of all surveys, the Doctorate Records File (DRF)—the survey database—contains public information on earlier doctorate recipients back to 1920. Thus, the DRF is a virtually complete data bank on more than 1.7 million doctorate recipients. The DRF also serves as the sampling frame for the biennial Survey of Doctorate Recipients (SDR), a longitudinal survey of science, engineering, and humanities doctorate recipients employed in the United States.

ANNUAL CENSUS OF NEW RESEARCH DOCTORATE RECIPIENTS:

SED collects self-reported data on:

- Demographic characteristics
- Educational history from high school to doctorate
- Mechanisms of financial support in graduate school
- Debt related to education
- Postgraduation plans

Purpose

To obtain consistent, annual data on individuals receiving research doctorates from U.S. institutions for the purpose of assessing trends in Ph.D. production.

Components

There is one component to the SED.

Survey of Earned Doctorates. The doctorate institution is responsible for administering the surveys to research doctoral candidates and, for the hard-copy version of the survey, collecting the completed questionnaires for mailback to the survey contractor. The doctorate recipients themselves complete the surveys. The following information is collected in the SED: all postsecondary degrees received and years awarded (although only the *first* baccalaureate, master's, first-professional, and doctoral degrees are entered in the database); years spent as a full-time student in graduate school; specialty field of doctorate; type of financial support received in graduate school; level of debt incurred in undergraduate and graduate school; employment/study status in the year preceding doctoral award; postgraduation plans (how definite, study vs. employment, type of employer, location, and basic annual salary); high school location and year of graduation;

demographic characteristics (sex, race/ethnicity, date and place of birth, citizenship status, country of citizenship for non-U.S. citizens, marital status, number of dependents, disability status, educational attainment of parents); and personal identifiers (name, last four digits of the Social Security Number, and permanent address). Dissertation field is keyed both as verbatim text and as a numeric code.

Periodicity

Annual since inception of the SED in the 1957–58 academic year. The database also includes basic information (obtained from public sources) on doctorates for the years 1920 to 1957.

2. USES OF DATA

The results from the SED are used by government agencies, academic institutions, and industry to address a variety of policy, education, and human resource issues. The survey is invaluable for assessing trends in doctorate production and the characteristics of Ph.D. recipients. The SED data are used to monitor the educational attainment of women and minorities, particularly in science and engineering. The increasing numbers of foreign citizens earning doctorates in the United States are studied by country of origin, field of concentration, sources of graduate school support, and the U.S. “stay” rate after graduation. Trends in time-to-doctorate are also analyzed by field, type of support received, and personal characteristics (such as marital status). Data on postdoctoral plans provide insight into the labor market for new Ph.D. recipients, whose careers can be followed in the longitudinal Survey of Doctorate Recipients, whose sample is drawn from the SED.

There is also substantial interest in the institutions attended by Ph.D. recipients. Doctorate-granting institutions frequently compare their survey results with peer institutions, and undergraduate institutions want to know their contribution to doctorate production. The availability of Carnegie Classifications in the DRF facilitates meaningful comparisons of the institutions attended by different demographic groups (e.g., men vs. women). Separate indicators for Historically Black Colleges and Universities (HBCUs) can allow researchers to examine the roles these institutions play in the educational attainment of Blacks.

3. KEY CONCEPTS

Some of the key terms and analytic variables in the SED are described below.

Research Doctorate. Any doctoral degree that (1) requires the completion of a dissertation or equivalent project of original work (e.g., musical composition); and (2) is not primarily intended as a degree for the practice of a profession. While the most typical research doctorate is the Ph.D., there are more than 20 other degree types (e.g., Ed.D., D.Sc., D.B.A.). Not included in this definition are professional doctorates: M.D., D.D.S., D.V.M., O.D., D.Pharm., Psy.D., J.D., and other similar degrees.

Doctorate-Granting Institution. Any postsecondary institution in the United States that awards research doctorates (as defined above) and is accredited at the higher education level by an agency recognized by the Secretary of the U.S. Department of Education. There are over 420 research doctorate-granting institutions.

Field of Doctorate. Specialty field of doctoral degree, as reported by the doctorate recipient. There are over 290 fields in the SED Specialties List, grouped under the following umbrellas: agricultural sciences/natural resources; biological/biomedical sciences; health sciences; engineering; computer and information sciences; mathematics; physical sciences (subdivided into astronomy, atmospheric science and meteorology, chemistry, geological and earth sciences, physics, and ocean/marine sciences); psychology; social sciences; humanities (subdivided into history, letters, foreign languages and literature, and other humanities); education (subdivided into research and administration, teacher education, teaching fields, and other education); and professional fields (subdivided into business management/administration, communication, and other professional fields). Because field of doctorate is designated by the doctorate recipient, the classification in the SED may differ from that reported by the institution in the NCES Integrated Postsecondary Education Data System (IPEDS) Completions Survey (see chapter 12).

Time-to-Doctorate. There are two standard, published measures of time-to-doctorate. The first measures the total elapsed time between bachelor’s degree receipt and doctorate degree receipt and can only be computed if baccalaureate year is known. *T*he second time-to-doctorate variable gauges the time between entry into graduate school (in any program or capacity, and in any university) and doctoral award. Both of these

measures are computed from items in the educational history section of the questionnaire.

Source of Support. Any source of financial support received during graduate school. Doctorate recipients are asked to mark all types of support received and to indicate the primary and secondary sources of support. For most SED years, sources are categorized as own/family resources; university related (teaching and research assistantships, university fellowships, college work-study); federal research assistantships (by agency); other federal support (by mechanism and agency); nonfederal U.S. nationally competitive fellowships (by funding organization); student loans (Stafford, Perkins); and other sources (business/employer, foreign government, state government).

In 1997–98, the number of source options was reduced from 35 to 13. Sources are no longer identified by the specific provider (e.g., federal agency, foundation, loan provider) since students do not always have that knowledge. Only the mechanism of support (e.g., fellowship, research assistantship, loan) is now requested. Most current categories are aggregates of multiple categories in previous questionnaires. For example, the new category "research assistantship" (RA) combines five earlier categories: university-related RA, NIH RA, NSF RA, USDA RA, and other federal RA. The following three categories are new as of 1997–98: grant, internship or clinical residency, and personal savings.

4. SURVEY DESIGN

Target Population

All individuals awarded research doctorates from accredited colleges and universities in the United States between July 1 of one year and June 30 of the following year. Currently, about 49,000 research doctorates are awarded annually by over 420 institutions located in the United States and Puerto Rico. Institutions in other U.S. jurisdictions do not grant research doctorates.

Sample Design

The SED is a census of all recipients of research doctorates in the United States and Puerto Rico.

Data Collection and Processing

The data collection and editing process spans a 21-month period ending 9 months after the last possible graduation date (i.e., June 30). The update of the database and preparation of tables for the first data release generally require another 4 to 6 months. From the inception of the SED in 1957–58 through the 1995–96 cycle, the survey was conducted by the National Research Council (NRC) of the National Academy of Sciences. In 1996–97, the SED was conducted by the NRC and processed by the new survey contractor, NORC. NORC has conducted all administrations since. The 1996–97 and 1997–98 administrations are considered a transition period. Not all NRC procedures were implemented in this period, and NORC continues to develop and test new procedures.

Reference Dates. The data are collected for an academic year, which includes all graduations from July 1 of one year through June 30 of the following year.

Data Collection. In advance of each administration of the survey, the contractor staff reviews the listings of accredited U.S. institutions in the Higher Education Directory to confirm that past participants are still doctorate granting and identify accredited institutions that are newly doctorate granting. As further confirmation of doctorate-granting status, the degree levels offered are checked in the IPEDS Institutional Characteristics data file (see chapter 12). By May of each year, questionnaires are mailed to the institutions for distribution to doctoral candidates who expect to receive their degree between July 1 of that year and June 30 of the following year. Institutional Coordinators are responsible for the distribution, collection, and return of the surveys. They are asked to provide official graduation lists or commencement programs along with the questionnaires and to provide addresses for students who did not complete questionnaires.

The vast majority of completed questionnaires (87 percent in 2008) are hard-copy versions of the SED survey instrument. A web-based SED option was implemented in 2001. Institutions distribute a link to the SED survey registration web page when students apply for graduation. Upon registering, students receive a PIN and password information via e-mail as well as the URL to the web survey instrument. This process enables coordinators to track the SED completion status of students who choose the web option. Utilization of the web option has grown over time, and accounted for 11 percent of the completed SED surveys in 2008. A third mode of data collection, an abbreviated questionnaire administered through computer-assisted telephone interviewing (CATI) that was initiated in 2005, accounted for the remaining 2 percent of completed surveys in 2008.

Upon receipt of a graduation batch, the contractor staff compares the names of students on completed questionnaires (“self-reports”) with the names in the commencement program or official graduation list. Any discrepancies are followed up with the institution for confirmation of graduation. If an address for a nonrespondent is provided by the institution or found through other means, a letter and questionnaire are mailed (or e-mailed) to the individual to request completion of the survey. A second mail/e-mail attempt is made to elicit participation if a response is not received within a month. Telephone solicitations using the CATI SED data collection mode follow the mail/e-mail efforts. In recent years, these follow-up efforts have yielded enough completed surveys to increase the survey’s overall self-report rate by 5 to 7 percentage points.

For doctorate recipients whose survey returns are still missing after these mailings, “skeleton” records are created from information contained in commencement programs or graduation lists: name; doctorate institution, field, and year; similar information for baccalaureate and master’s degrees; and sex (if it can be positively assumed from the name). Skeleton records have accounted for 7.3 to 8.8 percent of the records each year during the 2000s. In addition, a small percentage of surveys every year (usually less than 1 percent) are classified as “institutional” returns, having been completed by the institutions with whatever information was available to them. While institutional returns may contain more information than is available from commencement programs, their information is minimal compared to that in the self-reported surveys.

Survey contractor staff undergoes intensive training in the complexities of coding and checking procedures and is monitored throughout the collection cycle.

Data Processing. The SED processing includes two special efforts to increase response rates for key items. First, the data entry procedures used by both the NRC and NORC include triggers if any of eight “critical” items is missing: date of birth, sex, citizenship status, country of citizenship (if foreign), race/ethnicity, baccalaureate institution, baccalaureate year, and postdoctoral location. If any of these items is absent, a “missing information letter” (MIL) is generated and sent to the respondent. For these cases, five noncritical items (if missing) are also requested: birthplace, high school graduation year, high school location, master’s institution, and year of master’s degree.

Then, a second follow-up effort requests the same critical items from the doctorate-granting institutions, both for individuals who never completed a survey (skeletons) and for individuals who completed a survey (self-reports) but did not return the MIL. Because of the lower MIL yield during the transition period, more information was requested from institutions in 1996–97 and 1997–98. Respondents are now asked to provide the name and contact information of a person who is likely to know where they can be reached.

Editing. Records are processed through a multilayered edit routine that checks all variables for valid ranges of values and reviews the interrelationships among variables. The NRC performed these edits and the correction of errors online during data entry; then the full data file was processed a second time through selected edits after survey closure. NORC’s computer-assisted data entry (CADE) system also includes built-in range edits, but the interrelationship (consistency) edits are done after CADE is completed and after derived variables are created. There are more than 130 edit tests for the SED: about 20 range edits (all hard, mandatory edits that cannot be overridden) and nearly 120 interrelationship edits. About two-thirds of the interrelationship edits are hard edits. The remaining third are soft edits, which can be overridden after the responses are double-checked and verified as accurate.

The entire battery of edit tests was reviewed during the 1994–95 SED cycle. A large set of interrelationship tests was developed at this time to verify the accuracy of foreign-country coding for the various time frames covered in the survey. Other interrelationship tests check for reasonable time frames in the doctorate recipient’s chronology, from date of birth through date of doctoral award. Still others verify that the appropriate items are answered in a skip pattern (e.g., study vs. employment postdoctoral plans).

Estimation Methods

No weighting is performed since the SED is a census. Some logical assumptions are made during coding and updating of the database. For example, U.S. citizenship is assumed for Ph.D. recipients who designate their ethnicity as Puerto Rican since, legally, Puerto Ricans are U.S. citizens. Entries of “China” in country of citizenship may be recoded to either Taiwan or the People’s Republic of China, based on the locations of birthplace, high school, baccalaureate institution, and master’s degree institution. Postdoctoral plans are assumed to be employment if items in the employment section are answered and the postdoctoral study section is blank. Postdoctoral study is assumed if the opposite scenario is indicated.

Recent Changes

During the 1990s, the National Science Foundation asked the NRC to implement several new procedures in

an effort to improve both the quantity and quality of the SED data. Since the 1989–90 SED, there has been rigorous follow-up of complete nonrespondents and respondents who do not answer key data items. Race/ethnicity, postdoctoral location, and country of citizenship (if foreign) were first followed up in the 1989–90 cycle, increasing the completeness of these items from that time forward. In the mid-1990s, more than 100 new edit tests were implemented to check the coding of certain foreign countries for specific time frames. In recent years, the survey instrument has been reformatted a number of times to make it more respondent-friendly. Although the content has remained the same, the survey form was expanded from 4 to 12 pages in 1996, reduced to 8 pages in 2001, expanded to 10 pages in 2007, and expanded again to 12 pages in 2010.

During the 1996–97 cycle, the contract for conducting the SED was transferred from the NRC to NORC; this has brought some changes in procedures, as documented in earlier sections. In addition, the 1997–98 questionnaire included a major revision to the source of support question; the response set was changed from specific providers and mechanisms of support to only mechanisms. The marital status question was also changed in 1997–98 to (1) separate "widowed" from "separated/divorced" and (2) add a new category for "living in a marriage-like relationship."

Future Plans

Additional changes to the SED are under consideration, both to capture new data relevant to current issues in graduate education and to collect better data through existing questions.

5. DATA QUALITY AND COMPARABILITY

The 1990s brought a reexamination of all operational processes, introduction of state-of-the-art technologies, evaluations of data completeness and accuracy, and renewed efforts to attain even higher response rates for every item in the survey. A Technical Advisory Committee was established to guide the conduct of the SED with a look toward the future. A Validation Study was conducted to assess the limitations of the SED data, and data user groups were convened to advise on survey content. The survey instrument was reformatted to make it more respondent-friendly, and questions were revised in 2004 to collect more complete and accurate information. Beginning with the SED 2004, some Federal sponsor-approved changes were made to the standard questionnaire; questions were added to gather data on additional postsecondary degrees, master's degree as a prerequisite (formerly a check box and not a separate item), and postdoctoral position. In addition, the Education History items were redesigned and reformatted to ask only for information on completed degrees. Response codes for various items were also modified.

Sampling Error

The SED is a census and, thus, is not subject to sampling error.

Nonsampling Error

The main source of nonsampling error in the SED is measurement error. Coverage error is believed to be very limited. Unit and item response rates have been very high and relatively stable since the first survey in 1957–58 (although they were somewhat lower during the transfer of the SED administration to the new contractor).

Coverage Error. The SED is administered to a universe of research doctorate recipients identified by the universe of research doctorate-granting institutions. Therefore, undercoverage might result from (1) an incomplete institution universe; and/or (2) an incomplete enumeration of research doctorate recipients. The SED coverage has been evaluated and the uncoverage rate has been found to be less than 1 percent, due to the high visibility of doctorate-granting institutions and a comprehensive approach to data collection.

Every year, the universe of institutions is reviewed and compared to the institutional listings in the *Higher Education Directory* and other sources to determine the current list of doctorate-granting institutions. Any institutions newly determined to be doctorate granting are contacted for verification of doctorate-granting status and then invited to participate in the SED. A few qualifying institutions refuse to participate, but it is known from the IPEDS Completions Survey that these institutions contribute minimally to the overall doctorate population.

Individual doctorate recipients are enumerated through (1) survey forms completed by the new Ph.D. recipients and returned by the institution; (2) transmittal rosters that provide the official count of doctorates, the number of surveys completed and returned, and the names of individuals who did not complete surveys; and (3) commencement programs covering every graduation at an institution over the course of a year. Comparisons of the number of research doctorates in the SED with the total number of

doctorates reported by institutions in the IPEDS Completions Survey show that SED's coverage typically differs from IPED's by less than 1 percent.

Nonresponse Error. Targets have been set for both unit and item response in the SED. While the target rates are not always attained, response has been unusually high for a mail survey throughout the 40-plus years of the SED.

Unit Nonresponse. Basic information on non-respondents can be obtained from institutions or commencement programs, so records exist for all recipients of research doctorates. However, response to the SED is measured by the percentage of doctorate recipients who complete the surveys themselves (*self-report rate*), thus providing details that are not available from any other source. SED's goal is a stable self-report rate of 95 percent. This rate has been achieved or surpassed in all but 21 of the 51 surveys processed to date (through the 2008 SED). Response first fell below the target rate in 1986 and stayed low throughout the rest of the 1980s, at which time site visits and intensive follow-up procedures were initiated in an effort to increase the percentage of self-reported questionnaires. Response achieved the target level from 1990 to 1995 but has remained below target from 1996 to 2008 (ranging from 91.2 to 92.9 percent).

Because the SED is administered through doctorate-granting institutions, the self-report rate is dependent upon their overall cooperation and survey practices. Nonresponse tends to be concentrated in a small group of institutions. In the 2008 SED, 1 percent of the 421 doctorate-granting institutions accounted for 13 percent of the total nonrespondents, and the 19 percent of institutions with the highest nonresponse accounted for 65 percent of the total nonrespondents.

To improve tracking of institution response rates, NORC has devised an "early warning system" to identify institutions whose self-report rates lag behind the goal of 90 percent. Estimates for each seasonal graduation are developed based on the numbers for an institution's graduations in previous years. This system also allows monitoring of institutions with specific substantive interest for the SED (e.g., engineering schools, institutions awarding doctorates to large numbers of racial/ethnic minorities).

Item Nonresponse. Certain items are available for all doctorate recipients, whether or not they complete a questionnaire: name, doctorate institution, field of doctorate, month and year of doctoral award, and type of doctorate. This information is always provided by the institution in its commencement program or graduation list.

A 95 percent target is set for eight "critical" items: date of birth, sex, citizenship, country of citizenship (if foreign), race/ethnicity, baccalaureate institution, baccalaureate year, and postdoctoral location. From the 1989–90 SED (when rigorous follow-up of these items began) to the 1995–96 SED, all items but postdoctoral location achieved response rates above 95 percent. Rates for all critical items except sex and foreign country of citizenship fell below this goal in the 1996–97 and 1997–98 SED administrations, the transition period between contractors. In the 2008 administration, all of the critical items except sex achieved response rates below 95 percent.

Critical items are followed up through letters to self-reporting survey respondents and through requests to institutions for Ph.D. recipients who did not complete questionnaires. Thus, the response rates for these items often exceed the overall self-reporting rate for the survey. Because information can be obtained from sources other than the doctorate recipients, item response rates for the SED are computed on the universe of recipients, whether or not they responded to the survey.

Measurement Error. Most measurement error in the SED results from respondents' misinterpretation of questions or limited recall of past events. The 1994 Validation Study sought to determine the limitations of the SED data. Think-aloud interviews were conducted with recent Ph.D. recipients, who were asked to complete a second survey form within a few months of their original survey submission. The question on sources of support caused the most difficulty; few Ph.D. recipients responded exactly as they did in the initial survey. Problems with this item were confirmed by focus group discussions and comparisons of the SED results with raw data obtained from organizations that fund the various types of support. The source of support question was revised in the 1997–98 SED to request only the mechanism of support (e.g., research assistantship, fellowship, loan) rather than the actual source of funding (e.g., NSF, NIH), which some students do not know.

Interviewees were sometimes confused about the educational history section of the survey, thinking that information on short-term attendance at a school or attendance not leading to a degree was not required. Others were unsure about whether or not to include the time spent working on their dissertations. Such inconsistencies have an impact on time-to-doctorate computations. To address these issues, several new

questions on time to degree were added to the 2001 SED.

Several interviewees also had difficulty responding to the questions on postgraduation plans because, although they currently had a job, they wanted to indicate that they were still seeking a position that would satisfy their aspirations. These comments led to discussions among sponsors and other data users about the intent of the postdoctoral questions and what information is most relevant for policymaking.

Data Comparability

Because a prime use of the SED data is trend analysis, tremendous efforts have been made to maintain continuity of survey content. Five new items have been added since 2001: the basic annual salary for graduates with definite employment plans in the coming year, the level of tuition remission/waiver received during doctoral study, past enrollment in community college, master's degree as prerequisite for doctoral degree, and past or pending D.D.S. or M.D. degree. Occasional changes have been made to item response categories, sometimes affecting the comparability of data over time. For example, in 2001 the racial background question was changed to allow respondents to choose more than one option. In 2004 the education history questions were reformatted to ask specifically for information about the Ph.D., most recent master's degree, and first baccalaureate degree, and an additional question now asks about degrees earned beyond those three. For the items on disability status and debt level, format changes have occurred frequently enough to make comparisons with earlier years unreliable.

An additional modification was made to the 1997–98 questionnaire, affecting the sources of support item. The response set was overhauled to request information on only the mechanism of support (e.g., research assistantship, fellowship, loan) rather than mechanism *and* funder (e.g., NIH RA, NSF RA, university fellowship, NSF fellowship, Ford Foundation fellowship, Stafford loan, Perkins loan). As noted under Measurement Error above, focus groups and comparisons of the SED results with raw data obtained from organizations that fund the various types of support revealed that students do not always know the actual source of their support. The 1997–98 response set for the item on sources of support also includes three new categories: dissertation grant, internship/residency, and personal savings.

This major change has broken the time series for the sources of support item except for selected sources. NORC mapped the pre-1998 response categories to the new response set and then compared the 1997–98 distribution of responses to earlier distributions back to 1990. Significant shifts were observed in the proportions for some categories, raising concerns about whether the new code frame accurately captures the desired information on sources of support and suggesting the need for more cognitive work in this area. Therefore, users should be cautious about making generalizations regarding the financing of doctoral education over time.

Another comparability issue for the SED involves changes (generally, additions) made over the years to the survey's Specialties List, which is used to code fields for degrees, postdoctoral study, and employment. Because any specialties added to the list would have been coded into an "other" category (e.g., other biological sciences) in previous surveys, users should be careful in their interpretation of time-series field data at the most disaggregated level. The historical changes in the Specialties List are documented in *Science and Engineering Doctorates: 1960–91* (National Science Foundation 1993) and the subsequent series, *Science and Engineering Doctorate Awards* (Hill 2000).

While both unit and item response rates in the SED have been relatively stable through the years, fluctuations can affect data comparability. This is especially important to consider when analyzing data by citizenship and race/ethnicity, where very small fluctuations in response may result in increases or decreases in counts that do not reflect real trends. New procedures implemented in the early 1990s had a significant positive impact on response to these two items as well as to the items on foreign country of citizenship and postdoctoral location, making the data from 1990 to 1996 better in both quantity and quality than data from the late 1980s. Item response for citizenship and race/ethnicity has since fallen to the level of 1990 and earlier years, and item response for postdoctoral location is lower than in most years in the 1990s. Response to country of citizenship among non-U.S. citizens fell 3 percentage points (to 94.3 percent) in the first transition year (the 1997 SED) and has failed to return to pretransition levels.

The reformat of the questionnaire in 1995–96, described in earlier sections, resulted in substantial increases in response to primary source of support, postdoctoral work activity, and postdoctoral employment field. Users should take these changes into account when analyzing trends.

Comparisons with IPEDS. The IPEDS Completions Survey also collects data on doctoral degrees, but the

information is provided by institutions rather than by doctorate recipients. The number of doctorates reported in the IPEDS Completions Survey is slightly higher than in the SED. This difference is largely attributable to the inclusion in the IPEDS Completions Survey of nonresearch doctorates, primarily in the fields of theology and education. The differences in counts have been generally consistent since 1960, with ratios of IPEDS-to-SED counts ranging from 1.01 to 1.06. Because a respondent to the SED may not classify his or her specialty identically to the way the institution reports the field in the IPEDS Completions Survey, differences between the two surveys in the number of doctorates for a given field may be greater than the difference for all fields combined.

6. CONTACT INFORMATION

The National Science Foundation is the Systems Manager of Record for the Survey of Earned Doctorates. The micro-data can be used by institutions that enter into licensing agreements with NSF. The persons to contact concerning this are:

Mark K. Fiegener
Survey of Earned Doctorates
Division of Science Resources Statistics
National Science Foundation
Phone: (703) 292-4622
E-mail: mfiegene@nsf.gov

Stephen Cohen
Division of Science Resources Statistics
National Science Foundation
Phone: (703) 292-7769
E-mail: scohen@nsf.gov

For content information about the SED, contact:

NCES Contact:
Nancy Borkow
Phone: (202) 502-7311
E-mail: nancy.borkow@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

NSF Contact:
Mark K. Fiegener
Phone: (703) 292-4622
E-mail: mfiegene@nsf.gov

Mailing Address:
Human Resources Statistics Program
Division of Science Resources Statistics, Room 965
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

NORC Contact:
Vince Welch
Phone: (312) 759-4085
E-mail: welch-vince@norc.org

Mailing Address:
SED Project
National Opinion Research Center (NORC)
55 East Monroe Street, 20th Floor
Chicago, IL 60603

7. METHODOLOGY AND EVALUATION REPORTS

General
Shettle, C.F. (1995). *Guide to NSF Science and Engineering Resources Data* (NSF 95-318). Arlington, VA: National Science Foundation.

Survey Design
Dugoni, B., Lee, L., and Baldwin, A. (1999). *Report on Cognitive Research for the 2000 SED Questionnaire Development Task.* Chicago, IL: National Opinion Research Center.

Policy Research Methods. (1996). *Report on Cognitive Research for the 2000 SED Questionnaire Development Task*. Arlington, VA: Author.

Data Quality and Comparability
Hill, Susan T. (2000). *Science and Engineering Doctorate Awards: 1998* (NSF 00-304). Arlington, VA: National Science Foundation.

Ingram, L., and Ries, P. (1994). *Validation Study of the Survey of Earned Doctorates.* Washington, DC: National Research Council.

National Opinion Research Center. (1999). *Evaluation Report 1998: Quality Profile for the*

1997–1998 Survey of Earned Doctorates. Chicago, IL: Author.

National Research Council. (1997). *Evaluation Report 1996: Quality Profile for the 1995–1996 Survey of Earned Doctorates*. Washington, DC: Author.

National Science Foundation. (1993). *Science and Engineering Doctorates: 1960-91* (NSF 93-301). Washington, DC: Author.

Chapter 18: National Assessment of Educational Progress (NAEP)

1. OVERVIEW

The National Assessment of Educational Progress (NAEP) is mandated by Congress to assess the educational achievement of U.S. students and monitor changes in those achievements. As the only nationally representative and continuing assessment of what America's students know and can do in nine subject areas, NAEP serves as the "Nation's Report Card." The *main national* NAEP regularly assesses the achievement of students in grades 4, 8, and 12 at the national level. The *main state* NAEP assessed students at both grades 4 and 8 in at least one subject in 1990, 1992, 1994, 1996, 1998, 2000, 2002, and 2003. Since 2003, the main state NAEP has assessed students in at least two subjects, reading and mathematics, every 2 years at grades 4 and 8. The NAEP *Trial Urban District Assessment (TUDA)* assessed performance in selected large urban districts in 2002 in reading and writing at grades 4 and 8, and continued in 2003, 2005, 2007 and 2009 with reading and mathematics assessments at grades 4 and 8, and alternately science or writing.The *trend* NAEP tracks national long-term trends since the 1970s in mathematics and reading at ages 9, 13, and 17, and is given every 4 years. The national assessments were first implemented in 1969 and were conducted on an annual or biennial basis through 1995, and annually since 1996. The state assessments have been administered biennially since 1990.

In 1988, Congress established the National Assessment Governing Board (NAGB) to provide policy guidance for the execution of NAEP. The 26-member Governing Board is an independent, bipartisan group whose members include governors, state legislators, local and state school officials, educators, business representatives, and members of the general public. Its responsibilities include: select subject areas to be assessed; set appropriate student achievement levels; develop assessment objectives and test specifications; design the assessment methodology; and produce standards and procedures for interstate, regional, and national comparisons. NAEP is administered by the National Center for Education Statistics (NCES).

Purpose

To (1) monitor continuously the knowledge, skills, and performance of the nation's children and youth; and (2) provide objective data about student performance at the national, the regional, the state level (since 1990), and the district level (since 2002).

Components

NAEP comprises two unique assessments: *main and trend*; and there are three foci in the *main* assessment: *main national, main state* and *trial urban district*. Each of these assessments consists of four components: Elementary and Secondary School Students Survey; School Characteristics and Policies Survey; Teacher Survey; and

BIENNIAL SURVEY OF A SAMPLE OF ELEMENTARY/ SECONDARY STUDENTS

Two assessments:

- Main NAEP
- Trend NAEP

Three foci:

- Main National NAEP
- Main State NAEP
- Trial Urban District NAEP

Four component surveys:

- Elementary and Secondary School Students Survey
- School Characteristics and Policies Survey
- Teacher Survey
- SD/ELL Survey/ Excluded Student Survey

Students with Disabilities or English language learners(SD/ELL) Survey (for the *main national* NAEP) or Excluded Student Survey (for the *trend* NAEP).

In 1985, the Young Adult Literacy Study was also conducted nationally as part of NAEP, under a grant to the Educational Testing Service and Response Analysis Corporation; this study assessed the literacy skills of 21- to 25-year-olds. In addition, a High School Transcript Study (HSTS, see chapter 29) and a National Indian Education Study (NIES) are periodically conducted as components of NAEP.

Since 1996, the main national and state assessments have included accommodations for students with special needs.

National-level assessment. The *main national* NAEP and *trend* NAEP are both designed to report information for the nation and specific geographic regions of the country (Northeast, Southeast, Central, and West). However, these two assessments use separate samples of students from public and nonpublic schools: grade samples for the main national NAEP (grades 4, 8, and 12), and age/grade samples for trend NAEP (age 9/grade 4; age 13/grade 8; age 17/grade 11). The test instruments for the two assessments are based on different frameworks; the student and teacher background questionnaires vary; and the results for the two assessments are reported separately. (See "Elementary and Secondary School Students Survey" below for the subject areas assessed.)

The assessments in the *main national* NAEP follow the curriculum frameworks developed by NAGB and use the latest advances in assessment methodology. The test instruments are flexible so they can be adapted to changes in curricular and educational approaches. Recent assessment instruments for the main NAEP have been kept stable for short periods of time, allowing short-term trends to be reported from 1990 through 2009, except for the mathematics assessment for grade 12. In 2005, and 2009, NAGB introduced changes in the NAEP mathematics framework for grade 12 in both the assessment content and administration procedures.

To reliably measure change over longer periods of time, the *trend* NAEP must be used. For long-term trends, past procedures must be precisely replicated with each new assessment, and the survey instruments do not evolve with changes in curricula or educational practices. The instruments used today for the trend NAEP are relatively identical to those developed in the 1970s. Trend NAEP allows measurement of trends since 1971 in reading and 1973 in mathematics.

State-level assessments. The *main state* NAEP was implemented in 1990 on a trial basis and has been conducted biennially since that time. Participation of the states was completely voluntary until 2003. The reauthorization of the Elementary and Secondary Education Act, also referred to as the "No Child Left Behind Act," requires states that receive Title I funding to participate in state NAEP assessments in reading and mathematics at grades 4 and 8 every 2 years. State participation in other state NAEP subjects (i.e., science and writing) remains voluntary. Separate representative samples of students are selected for each jurisdiction to provide that jurisdiction with reliable state-level data concerning the achievement of its students. The state assessment included nonpublic schools in 1994, 1996, and 1998. This practice ended because of low participation rates. (See below for the subject areas assessed.)

The Trial Urban District Assessment. The *Trial Urban District Assessment (TUDA)* began assessing performance in selected large urban districts in 2002 in reading and writing; it continued in 2003 with reading and mathematics; in 2005 with reading, mathematics, and science; in 2007 with reading, mathematics and writing, and in 2009 with reading, mathematics and science. The program retains its trial status. The first TUDA occurred in reading and writing in 2002 for five urban districts. In 2003, nine districts were assessed in mathematics and reading. In 2005 and 2007, ten urban school districts participated in TUDA. The results for these districts are for public school students only. Results for District of Columbia public school students, normally included with NAEP's state assessment results, are also reported in TUDA in 2005 and 2007 in reading and mathematics. (Due to an insufficient sample size, the District of Columbia did not participate in the science assessment in 2005 and 2009 and the writing assessment in 2007.) Beginning in 2009, the TUDA results include only those charter schools that the district is accountable for.) Results for these districts are also compared with results for public school students in large central cities and the nation.

Elementary and Secondary School Students Survey. The primary data collected by NAEP relate to student performance and educational experience as reported by students. Major assessment areas include: reading, writing, mathematics, science, civics, U.S. history, geography, economics, and the arts.

Subjects assessed in the main national NAEP. In 1988, the *main national* NAEP assessed student performance in reading, writing, civics, and U.S. history, and conducted small assessments in geography and document literacy. In 1990, it assessed mathematics, reading, and science; in 1992, reading, mathematics, and writing; in 1994, reading, U.S. history, and world geography; and in 1996, science and mathematics. A probe of student performance in the arts at grade 8 was conducted in 1997. Reading, writing, and civics were assessed in 1998. (*Trend* NAEP was conducted in 1999.) In 2000, the main national NAEP assessed mathematics and science (and, for 4th-graders only, reading). In 2001, history and geography were assessed; in 2002, reading and writing. In 2003, the assessments were in reading and mathematics for 4th- and 8th-graders. In 2004, the main national NAEP assessed foreign language for 12th grade. In 2005, the assessments were in reading, mathematics, and science, and in 2006, in U.S. history and civics (and, for 12th-graders only, in economics). In 2007, reading and mathematics were assessed at grades 4 and 8, and writing at grades 8 and 12. In 2008, the arts were assessed at grade 8. In 2009, reading, mathematics and science were assessed at grade 4, 8 and 12. In 2010, U.S. history, civics and geography were assessed at grade 4, 8 and 12.

Subjects assessed in trend NAEP. The subjects assessed in *trend* NAEP are mathematics and reading (and, until 1999, writing; and, until 2004, science). The biennial assessments from 1988 through 1996 covered all subjects. Since 2004, the trend assessments have been scheduled to be administered in mathematics and reading every 4 years. The latest trend assessment was conducted in 2008, and the report was released in the spring of 2009.

Subjects assessed in the main state NAEP. Data representative of states were collected for the first time in the 1990 trial state assessment, when 8th-grade students were assessed in mathematics. In 1992, state-level data were collected in 4th-grade reading and mathematics, and in 8th-grade mathematics. In 1994, 4th-grade reading was assessed. In 1996, 4th-grade mathematics and 8th-grade mathematics and science were assessed. The 1998 NAEP collected state-level data in reading at grades 4 and 8, and writing at grade 8. The 2000 NAEP assessments covered mathematics and science, the 2002 assessments covered reading and writing, the 2003 assessments covered reading and mathematics, and the 2005 assessment covered reading, mathematics, and science. The 2007 state assessment covered reading and mathematics (and, for grade 8 only, writing). The 2009 state assessment covered reading, mathematics and science at grades 4 and 8, and reading and mathematics at grade 12.

Subjects assessed at TUDA. Data representative of urban districts were collected for the first time on a trial basis in selected large urban districts in 2002 with reading and writing assessments. In 2003, district-level data were collected in 4th- and 8th-grade reading and mathematics. In 2005, 4th- and 8th-grade reading, mathematics and science were assessed. In 2007, 4th- and 8th-grade reading, mathematics and writing were assessed. TUDA retains its trial status, and is scheduled for 2009 .

Student background questions. The student survey also asks questions about the student’s background, as well as questions related to the subject area and the student’s motivation in completing the assessment. Student background questions gather information about race/ethnicity, school attendance, academic expectations, and factors believed to influence academic performance, such as homework habits, the language spoken in the home, and the quantity of reading materials in the home. Some of these questions document changes that occur over time: these questions remain unchanged over assessment years.

Student subject-area questions. These questions gather three categories of information: time spent studying the subject, instructional experiences in the subject, and perceptions about the subject. Because these questions are specific to each subject area, they can probe in some detail the use of specialized resources (such as the use of calculators in mathematics classes).

Students are also asked how often they have been asked to write long answers to questions on tests or assignments that involve the tested subject. Before 2004, students were also asked how many questions they thought they answered correctly, how difficult they found the assessment, how hard they tried on this test compared to how hard they had tried on other tests or assignments they had taken that year in school, and how important it was to them to do well on this test.

School Characteristics and Policies Survey. This survey collects supplemental data about school characteristics and school policies that can be used analytically to provide context for student performance issues. Data are collected on enrollment, absenteeism, dropout rates, curricula,

testing practices, length of school day and year, school administrative practices, school conditions and facilities, size and composition of teaching staff, tracking policies, schoolwide programs and problems, availability of resources, policies for parental involvement, special services, and community services.

Teacher Questionnaire. This study collects supplemental data from teachers whose students are respondents to the assessment surveys. The first part of the teacher questionnaire tends to cover background and general training, and includes items concerning years of teaching experience, certifications, degrees, major and minor fields of study, course work in education, course work in specific subject areas, the amount of in-service training, the extent of control over instructional issues, and the availability of resources for the classroom. Subsequent parts of the teacher questionnaire tend to cover training in the subject area, classroom instructional information, and teacher exposure to issues related to the subject and the teaching of the subject. They also ask about pre- and in-service training, the ability level of the students in the class, the length of homework assignments, use of particular resources, and how students are assigned to particular classes.

SD/ELL Survey. This survey is completed in the *main* NAEP assessments (and the *trend* NAEP since 2004) by teachers of students who are selected to participate in NAEP but who are classified as either having disabilities (SD) or English language learners (ELL). Information is collected on the background and characteristics of each SD/ELL student and the reason for the SD/ELL classification, as well as on whether these students receive accommodations in district or statewide tests. For SD students, questions ask about the student's functional grade levels and special education programs. For ELL students, questions ask about the student's native language, time spent in special language programs, and level of English language proficiency. This survey is used to determine whether the student should take the NAEP assessment. If any doubt exists about a student's ability to participate in the assessment, the student is included. Beginning with the 1996 assessments, NAEP has allowed accommodations for both SD and ELL students.

Excluded Student Survey. This survey is completed in trend NAEP for students who are sampled for the assessment, but who are excluded by the school from participating in it. Following exclusion criteria used in previous trend assessments, a school can exclude students with limited English-speaking ability, students who are educable mentally retarded, and students who are functionally disabled—if the school judges that these students are unable to "participate meaningfully" in the assessment. This survey is only completed for those students who are actually excluded from the assessment (whereas the SD/ELL Survey in the main assessment is also completed for participating students who are SD or ELL students—see above).

High School Transcript Study. Transcript studies have been conducted in 1987, 1990, 1994, 1998, 2000, 2005, and 2009. The studies collect information on current course offerings and course-taking patterns in the nation's schools. Transcript data can be used to show course-taking patterns across years that may be associated with proficiency in subjects assessed by NAEP. Transcripts are collected from grade 12 students in selected schools in the NAEP sample. (For more information on the High School Transcript Studies, see chapter 29.)

National Indian Education Study. The National Indian Education Study (NIES) is a two-part study designed to describe the condition of education for American Indian and Alaska Native (AI/AN) students in the United States. The study is conducted by NCES on behalf of the U.S. Department of Education, Office of Indian Education. NIES is authorized under Executive Order 13336, "American Indian and Alaska Native Education", which was signed in 2004 to improve education efforts for AI/AN students nationwide.

Part I of NIES is conducted through NAEP and provides in-depth information on the academic performance of 4th– and 8th–grade AI/AN students in reading and mathematics. Part II of NIES is a survey that describes the educational experiences of the 4th– and 8th–grade AI/AN students who participated in the NAEP assessments. The survey focuses on the integration of native language and culture into school and classroom activities. Part II collects information through questionnaires for students, teachers, and principals.

Oral Reading Study. In 2002, NAEP conducted a special study on oral reading. The NAEP 2002 Oral Reading Study looked at how well the nation's 4th-graders can read aloud a grade-appropriate story. NAEP assessed a random sample of 4th-grade students selected for the NAEP 2002 reading and writing assessments. The assessment provided information about a student's fluency in reading aloud and examined the relationship between oral reading

accuracy, rate (or speed), fluency, and reading comprehension.

Technology-Based Assessment (TBA) Project. TBA was a NAEP project in 2000 to 2003.TBA was designed with five components—three empirical studies (Mathematics Online, Writing Online, and Problem Solving in Technology-Rich Environment); a conceptual paper (Computerized Adaptive Testing); and an online school and teacher questionnaire segment. The three empirical studies were the primary focus of the TBA Project and are discussed below.

The primary goals of the Mathematics Online (MOL) study were to understand how computer delivery affects the measurement of NAEP math skills, to gain insight into the operational and logistical mechanics of computer-delivered assessments, and to evaluate the ability of 4^{th}- and 8^{th}-graders to deal with mathematics assessments delivered on computer. At grade 8, an additional goal was to investigate the technical feasibility of generating alternate versions of multiple-choice and constructed-response items using “on-the-fly” technology. MOL was field tested in 2002.

The Writing Online (WOL) study was intended to help NAEP learn how computer delivery affects the measurement of NAEP performance-based writing skills, to gain insights into the operational and logistical mechanics of computer-delivered writing assessments, and to evaluate the ability of 8^{th}-graders to deal with writing assessments delivered on computer. WOL was field tested in 2002.

The Problem Solving in Technology-Rich Environments (TRE) study was designed to develop an example set of modules to assess problem solving using technology. These example modules use the computer to present multimedia tasks that cannot be delivered through conventional paper-and-pencil assessments, but which tap important emerging skills. TRE was field tested in 2003.

Charter School Pilot Study. NAEP conducted a pilot study of America’s charter schools and their students as part of the 2003 NAEP assessments in reading and mathematics at the 4^{th}-grade level. Charter schools are public schools of choice. They serve as alternatives to the regular public schools to which students are assigned. While there are many similarities between charter schools and other public schools, they do differ in some important ways, including the makeup of the student population and their location.

Student Achievement in Private Schools. To better understand the performance of students in private schools, NAEP performed two studies and has released a two-part series of reports. In the first report *Student Achievement in Private Schools: Results from NAEP 2000–2005 (*Perie, Vanneman, & Goldstein 2005), the results of the 2000, 2002, 2003, and 2005 assessments for all private schools and for the largest private school categories—Catholic, Lutheran, and conservative Christian—were compared with the results for public schools (where applicable). This report focused on important demographic differences between students nationwide in private and public schools. The goal of the second report (*Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling* [Braun, Jenkins, & Grigg 2006]) was to examine differences in mean NAEP reading and mathematics scores in 2003 between public and private schools when selected characteristics of students and/or schools were taken into account. Hierarchical linear models were employed to carry out the desired adjustments.

Periodicity

Annual from 1969 to 1979, biennial in even-numbered years from 1980 to 1996, after which it was annual. A probe of 8^{th}-graders in the arts was conducted in 1997 and again in 2008. State-level assessments, initiated in 1990, follow the same schedule as the main national assessments. Prior to 1990, NAEP was required to assess reading, mathematics, and writing at least once every 5 years. The previous legislation required assessments in reading and mathematics at least every 2 years, in science and writing at least every 4 years, and in history or geography and other subjects selected by the NAGB at least every 6 years.

The No Child Left Behind Act requires NAEP to conduct national and state assessments at least once every 2 years in reading and mathematics in grades 4 and 8. In addition, NAEP has conducted a national assessment in reading and mathematics in grade 12 every 4 years starting since 2005. TUDA began assessing performance in selected large urban districts in 2002 with reading and writing assessments and continued in 2003, 2005, 2007 and 2009 with reading and mathematics assessments. TUDA is scheduled for 2011 as well. The program retains its trial status. Finally, to the extent that time and money allow, NAEP will be conducted in grades 4, 8, and 12 at regularly scheduled intervals in additional subjects including writing, science, history, geography, civics, economics, foreign languages, and the arts.

NIES was conducted for the first time in 2005 as a part of NAEP, in accordance with Title VII, Part A of the Elementary and Secondary Education Act, 2001. The second NIES data collection took place in 2007, and

the third collection took place in 2009: NCES is planning to conduct NIES again in 2011.

2. USES OF DATA

NAEP is the only ongoing, comparable, and representative assessment of what American students know and can do in nine subject areas. Policymakers are keenly interested in NAEP results because they address national outcomes of education, specifically, the level of educational achievement. In addition, state-level and urban district-level data, available for many states since 1990 and for selected large urban districts since 2002, allow both state-to-state and district-to-district comparisons, and comparisons of individual states with the nation as a whole (as well as comparisons of urban districts with large central cities and the nation).

During NAEP's history, a number of reports across various subject areas have provided a wealth of information on students' academic performance, learning strategies, and classroom experiences. Together with the performance results, the basic descriptive information collected about students, teachers, administrators, and communities can be used to address the following educational policy issues:

- *Instructional practices*. What instructional methods are being used?
- *Students-at-risk*. How many students appear to be at-risk in terms of achievement, and what are their characteristics? What gaps exist between at-risk categories of students and others?
- *Teacher workforce*. What are the characteristics of teachers of various subjects?
- *Education reform*. What policy changes are being made by our nation's schools?

However, *users should be cautious in their interpretation of NAEP results. While NAEP scales make it possible to examine relationships between students' performance and various background factors, the relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables.* NAEP results are most useful when they are considered in combination with other knowledge about the student population and the education system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

NAEP materials such as frameworks and released questions also have many uses in the educational community. Frameworks present and explain what experts in a particular subject area consider important. Several states have used NAEP frameworks to revise their curricula. After most assessments, NCES publicly releases nearly one-third of the questions. Released constructed-response questions and their corresponding scoring guides have served as models of innovative assessment practices in the classroom.

3. KEY CONCEPTS

The achievement levels for NAEP assessments are defined below. For subject-specific definitions of achievement levels and additional terms, refer to NAEP technical reports, "report card" reports, and other publications.

Achievement levels. Starting with the 1990 NAEP, NAGB developed achievement levels for each subject at each grade level to measure how well students' actual achievement matches the achievement desired of them. The three levels are as follows:

- *Basic*. Partial mastery of the prerequisite knowledge and skills that are fundamental for proficient work at each grade.
- *Proficient*. Solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
- *Advanced*. This level signifies superior performance, and is attained by only a very small percentage of students (3–6 percent) at any of the three grade levels assessed.

4. SURVEY DESIGN

Target Population

Students enrolled in public and nonpublic schools in the 50 states and the District of Columbia who are deemed assessable by their school and classified in

defined grade/ age groups—grades 4, 8, and 12 for the *main national* assessments and ages 9, 13, and 17 for the *trend* assessments in science, mathematics, and reading. Grades 4 and/or 8 are usually assessed in the *state* assessments and *TUDA*; the number of grades assessed has varied in the past, depending on the availability of funding (although testing for 4^{th}- and 8^{th}-graders in reading and mathematics every 2 years is now required for states that receive Title I funds). Only public schools were included in the *state* NAEP prior to 1994 and after 1998. Only public schools are included in *TUDA*.

Sample Design

For the national assessments, probability samples of schools and students are selected to represent the diverse student population in the United States. The numbers of schools and students vary from cycle to cycle, depending on the number of subjects and items to be assessed. A national sample will have sufficient schools and students to yield data for public schools and each of the four NAEP regions of the country, as well as sex, race, degree of urbanization of school location, parent education, and participation in the National School Lunch Program. A separate grade 12 sample of schools is also selected to produce national and regional estimates, as state NAEP does not yet include grade 12 (a pilot study of grade 12 state NAEP was conducted in 2009). A national sample of nonpublic (private) schools is also selected for grades 4, 8, and 12. This sample is designed to produce national and regional estimates of student performance for private schools.

In the state assessment, a sample of schools and students is selected to represent a participating state. In a state, on average 2,500 students in approximately 100 public schools are selected per grade, per subject assessed. The selection of schools is random within classes of schools with similar characteristics; however, some schools or groups of schools (districts) can be selected for each assessment cycle if they are unique in the state. For instance, a particular district may be selected more often if it is located in the state's only major metropolitan area or has the majority of the state's Black, Hispanic, or other race/ethnicity population. Additionally, even if a state decides not to participate at the state level, schools in that state identified for the national sample will be asked to participate.

Typically, 30 students per subject per grade are selected randomly in each school. Some of the students who are randomly selected are classified as SD or ELL. NAEP's goal is to assess all students in the sample, and this is done if at all possible.

NAEP's multistage sampling process involves the following steps:

- selection of schools (public and nonpublic) within strata and

- selection of students within the selected schools.

Selection of schools. In this stage of sampling, public schools in each state (—including Bureau of Indian Education [BIE] schools and Department of Defense Education Activity [DoDEA] schools)—and nonpublic schools in each state (including Catholic schools) are listed according to the grades associated with the three age classes: age class 9 refers to age 9 or grade 4 in the trend NAEP (or grade 4 in the main NAEP); age class 13 refers to age 13 or grade 8 in the trend NAEP (or grade 8 in the main NAEP); age class 17 refers to age 17 or grade 11 in the trend NAEP (or grade 12 in the main NAEP).

The school lists are obtained from two sources. Regular public, BIE, and DoDEA schools are obtained from the school list maintained by Common Core of Data. (See chapter 2). Catholic and other nonpublic schools are obtained from the NCES Private School Universe Survey (PSS). (See chapter 3.) To ensure that the state samples provide an accurate representation, public schools are stratified by urbanization, enrollment of Black, Hispanic, or other race/ethnicity students, and median house-hold income. Nonpublic schools are stratified by type of control (e.g., parochial, nonreligious), urban status, and enrollment per grade. Once the stratification is completed, the schools within each state are assigned a probability of selection that is proportional to the number of students per grade in each school.

Prior to 2005, DoDEA overseas and domestic schools were reported separately. In the 2005 assessments, all DoDEA schools, both domestic and overseas, were combined into one jurisdiction. In addition, the definition of the national sample changed in 2005; it now includes all of the overseas DoDEA schools.

The manner of sampling schools for the long-term trend assessments is very similar to that used for the main assessments. The primary difference is that in long-term trend nonpublic schools and schools with high enrollment of Black, Hispanic, or other race/ethnicity students are not oversampled. Schools

are not selected for both main and long-term trend assessments at the same age/grade. The long-term trend assessments use a nationally representative sample and do not report results by state.

Selection of students. This stage of sampling involves random selection of national samples representing the entire population of U.S. students in grades 4, 8, and 12 for the main assessment and the entire population of students at ages 9, 13, and 17 for the long-term trend assessment. Typically, 30 students per subject per grade are selected randomly in each school. Some of the students who are randomly selected are classified as SD or ELL. A small number of students selected for participation are excluded because of limited English proficiency or severe disability.

To facilitate the sampling of students, a consolidated list is prepared for each school of all age-eligible students (long-term trend assessments) or all grade-eligible students (main assessments) for the age class for which the school is selected. A systematic selection of eligible students is made from this list—unless all students are to be assessed—to provide the target sample size.

For each age class (separately for long-term trend and main samples), maxima are established as to the number of students who are to be selected for a given school. In those schools that, according to information in the sampling frame, have fewer eligible students than the established maxima, each eligible student enrolled at the school is selected in the sample. In other schools, a sample of students is drawn. The maximum sample sizes are established in terms of the number of grade-eligible students for the main samples, and in terms of the number of students in each age class for the trend samples.

Excluded students Some students are excluded from the student sample because they are deemed unassessable by school authorities. The exclusion criteria for the main samples differ somewhat from those used for the long-term trend samples. In order to identify students who should be excluded from the main assessments, school staff members are asked to identify those SD or ELL students who do not meet the NAEP inclusion criteria. School personnel are asked to complete an SD/ELL questionnaire for all SD and ELL students selected into the NAEP sample, whether they participate in the assessment or not. Prior to 2004, for the long-term trend assessments, excluded students were identified for each age class, and an Excluded Student Survey was completed for each excluded student. Beginning in 2004, both trend and main NAEP assessments use identical procedures.

For the special study of Students with Disabilities or Limited English Proficient (SD/LEP) inclusion in the 1996 main assessment, oversampling procedures were applied to SD/LEP students at all three grades in sample types 2 (accommodations not allowed) and 3 (accommodations allowed) for mathematics and in sample type 3 for science. (Sample type denotes whether or not a session may allow such accommodations.)

Main national NAEP sample sizes. Not all subject areas are assessed in every assessment year. In 2009, the main national NAEP assessed students in reading, mathematics and science at grades 4, 8 and 12. For the main national NAEP, a nationally representative sample of more than 350,000 students at grades 4, 8, and 12 participated in these assessments. The main national math assessment sampled 168,800 4th grade students, 161,700 8th grade students, and 48,900 12th grade students; the reading assessment sampled 178,800 4th grade students, 160,900 8th grade students, and 51,700 12th grade students. The science assessment sampled 156,500 4th grade students, 151,100 8th grade students, and 11,100 12th grade student. The mathematics, reading, and science assessments were conducted in the same 9,600 4th grade schools, 7,110 8th grade schools, and 1,680 12th grade schools.

TUDA sample sizes. In 2009, eighteen urban districts (including District of Columbia) participated in TUDA in math and reading and 17 urban districts participated in TUDA in science. The sample design for TUDA districts provides for oversampling. For the five largest TUDA districts—New York City, Los Angeles, Chicago, Miami and Houston—the target student sample sizes are three-quarters the normal size of the state sample. For the other twelve districts (Atlanta, Austin, Baltimore City, Boston, Charlotte, Cleveland, Detroit, Fresno, Jefferson County, KY, Milwaukee, Philadelphia, and San Diego), the target student sample sizes are half the normal size of the state sample. The larger samples allow reliable reporting about subgroups in these districts.

Students in the TUDA samples are considered part of the state and national samples. For example, the data for students tested in the Chicago sample will be used to report results for Chicago, but will also contribute to Illinois' estimates (and, with appropriate weights, to national estimates). Chicago has approximately 20 percent of the students in Illinois; therefore Chicago

will contribute 20 percent, and the rest of the state will contribute 80 percent, to Illinois' results.

Long-term trend NAEP sample sizes. The long-term trend assessment tested the same four subjects across years through 1999, using relatively small national samples. Samples of students were selected by age (9, 13, and 17) for mathematics, science, and reading, and by grade (4, 8, and 11) for writing. Students within schools were randomly assigned to either mathematics/science or reading/writing assessment sessions subsequent to their selection for participation in the assessments. In 2004, science and writing were removed from the trend assessments; the trend assessments are now scheduled to be administered in mathematics and reading every 4 years (but not in the same years as the main assessments). In 2004, approximately 24,100 students took the modified[1] reading assessment, while about 14,000 took the bridge[2] reading assessment. In 2004, approximately 22,400 students took the modified mathematics assessment, while about 14,700 took the bridge mathematics assessment. The latest trend assessment was conducted in 2008, with approximately 26,600 students assessed in reading, and 26,700 students assessed in mathematics.

NIES Part II sample sizes. The NIES Part II sample is designed to produce information representative of the target population of all fourth- and eighth-grade AI/AN students in the United States. In 2005, the sample included about 5,600 eligible students at approximately 550 schools located throughout the United States. The sample consisted of approximately 84 percent public, 4 percent private, and 12 percent BIE schools (unweighted). In 2007, the NIES Part II sample included about 12,900 AI/AN students at approximately 1,900 schools at grade 4 and 14,600 AI/AN students at 2,000 schools at grade 8 located throughout the United States. The sample consisted of approximately 94 percent public, 1 percent private, and 5 to 6 percent BIE schools at grades 4 and 8 (as well as a small number of DoDEA schools). All BIE schools were part of the sample. In 2009, the NIES Part II sample consisted of about 12,300 grade 4 students in approximately 2,300 schools and approximately 10,400 students in grade 8 at about 1,900 schools.

Assessment Design

Since 1988, NAGB has selected the subjects for the main NAEP assessments. NAGB also oversees the creation of the frameworks that underlie the assessments and the specifications that guide the development of the assessment instruments.

Development of framework and questions. NAGB uses an organizing framework for each subject to specify the content that will be assessed. This framework is the blueprint that guides the development of the assessment instrument. The framework for each subject area is determined with input from teachers, curriculum specialists, subject-matter specialists, school administrators, parents, and members of the general public.

Unlike earlier multiple-choice instruments, current instruments dedicate a majority of testing time to constructed-response questions that require students to compose written answers. Constructed-response questions provide a separate means of assessing ability that taps recall, not recognition.

The questions and tasks in an assessment are based on the subject-specific frameworks. They are developed by teachers, subject-matter specialists, and testing experts under the direction of NCES and its contractors. For each subject-area assessment, a national committee of experts provides guidance and reviews the questions to ensure that they meet the framework specifications. For each state-level assessment, state curriculum and testing directors review the questions that will be included in the NAEP state component.

Matrix sampling. Several hundred questions are typically needed to reliably test the many specifications of the complex frameworks that guide NAEP assessments. However, administering the entire collection of cognitive questions to each student would be far too time consuming to be practical. Matrix sampling allows the assessment of an entire subject area within a reasonable amount of testing time, in most cases 50 minutes. By this method, different portions from the entire pool of cognitive questions are printed in separate booklets and administered to different but equivalent samples of students.

The type of matrix sampling used by NAEP is called focused, balanced incomplete block (BIB) spiraling. The NAEP BIB design varies according to subject area.

Data Collection and Processing

Since 1983, NCES has conducted NAEP through a series of contracts, grants, and cooperative agreements with the Educational Testing Service

[1] The modified assessment included new items and features, representing the new design.
[2] The bridge assessment replicates the assessment given in the previous assessment year.

(ETS) and other contractors. ETS is directly responsible for developing the assessment instruments, analyzing the data, and reporting the results. Westat selects the school and student samples, trains assessment administrators, and manages field operations (including assessment administration and data collection activities). NCS Pearson is responsible for printing and distributing the assessment materials and for scanning and scoring students' responses.

Reference dates. Data for the main national NAEP and main state NAEP are collected from the last week in January through the first week in March. Data for the long-term trend NAEP are collected during the fall for age 13; during the winter of the same school year for age 9; and during the spring for age 17.

Data collection. Before 2002, NCES had relied heavily on school administrators for the conduct of main state NAEP assessments. Beginning with the 2002 assessments, however, NAEP contractor staff has conducted all NAEP assessment sessions. Obtaining the cooperation of the selected schools requires substantial time and energy, involving a series of mailings that includes letters to the chief state school officers and district superintendents to notify the sampled schools of their selection; additional mailings of informational materials; and introductory in-person meetings where procedures are explained.

The questionnaires for the School Characteristics and Policies Survey, the Teacher Survey, and the SD/ELL Survey are sent to schools ahead of the assessment date so that they can be collected when the assessment is administered. Questionnaires not ready at this time are retrieved later, either through a return visit by NAEP personnel or through the mail.

NCS Pearson produces the materials needed for NAEP assessments. NCS Pearson prints identifying barcodes and numbers for the booklets and questionnaires, pre-assigns the booklets to testing sessions, and prints the booklet numbers on the administration schedule. These activities improve the accuracy of data collection and assist with the BIB spiraled distribution process.

Assessment exercises are administered either to individuals or to small groups of students by specially trained field personnel. For all three ages in the long-term trend NAEP, the mathematics questions administered using a paced audiotape before 2004. Since 2004, the long-term trend assessments have been administered through test booklets read by the students.

For the long-term trend assessments, Westat hires and trains approximately 85 field staff to collect the data. For the 2009 main national and state assessments, Westat hired and trained about 7,000 field staff to conduct the assessments.

After each session, Westat staff interview the assessment administrators to receive their comments and recommendations. As a final quality control step, a debriefing meeting is held with the state supervisors to receive feedback that will help improve procedures, documentation, and training for future assessments.

For NIES Part II, NCES data collection contractor staff visit the schools to administer survey questionnaires. Students complete the questionnaires in group settings proctored by study representatives. In order to decrease the possibility that survey responses might be adversely affected by students' reading levels, the questions are read aloud to all grade 4 students and to grade 8 students who school staff think might need assistance. In addition, the study representatives are available to answer any questions that students have as they work on the questionnaires.

In 2005, survey materials were mailed to about 20 percent (unweighted) of the NIES Part II schools (primarily schools that were remotely located and had only a few AI/AN students), and the schools were asked to administer the questionnaires and return them by mail. Detailed instructions were provided for identifying teachers and students to be surveyed, administering the student questionnaires, responding to questions from students, and labeling and returning survey materials. Although the mail mode was used at about 20 percent (unweighted) of the sampled schools, these schools generally had only one or two sampled students. Thus, only about 2 percent of the sampled students were at mail-mode schools. The mail-mode data collection procedure was discontinued after the 2005 administration of NIES Part II.

Data processing. NCS Pearson handles all receipt control, data preparation and processing, scanning, and scoring activities for NAEP. Using an optical scanning machine, NCS Pearson staff scans the multiple-choice selections, the handwritten student responses, and other data provided by students, teachers, and administrators. An intelligent data entry system is used for resolution of the scanned data, the entry of documents rejected by the scanning machine, and the entry of information from the

questionnaires. An image-based scoring system introduced in 1994 virtually eliminates paper handling during the scoring process. This system also permits online monitoring of scoring reliability and creation of recalibration sets.

ETS and NCS Pearson develop focused, explicit scoring guides with defined criteria that match the criteria emphasized in the assessment frameworks. The scoring guides are reviewed by subject-area and measurement specialists, the instrument development committees, NCES, and NAGB to ensure consistency with both question wording and assessment framework criteria. Training materials for scorers include examples of student responses from the actual assessment for each performance level specified in the guides. These exemplars help scorers interpret the scoring guides consistently, thereby ensuring the accurate and reliable scoring of diverse responses.

The image-based scoring system allows scorers to assess and score student responses online. This is accomplished by first scanning the student response booklets, digitizing the constructed responses, and storing the images for presentation on a large computer monitor. The range of possible scores for an item also appears on the display; scorers click on the appropriate button for quick and accurate scoring. The image-based scoring system facilitates the training and scoring process by electronically distributing responses to the appropriate scorers and by allowing ETS and NCS Pearson staff to monitor scorer activities consistently, identify problems as they occur, and implement solutions expeditiously. The system also allows the creation of calibration sets that can be used to prevent drift in the scores as signed to questions. This is especially useful when scoring large numbers of responses to a question (e.g., more than 30,000 responses per question in the state NAEP). In addition, the image-based scoring system allows all responses to a particular exercise to be scored continuously until the item is finished, thereby improving the validity and reliability of scorer judgments.

The reliability of scoring is monitored during the coding process through (1) backreading, where table leaders review about 10 percent of each scorer's work to confirm a consistent application of scoring criteria across a large number of responses and across time; (2) daily calibration exercises to reinforce the scoring criteria after breaks of more than 15 minutes; and (3) a second scoring of 25 percent of the items appearing only in the main national assessment and 6 percent of the items appearing in both the main national and state assessments (and a comparison of the two scores to give a measure of interscorer reliability). To monitor agreement across years, a random sample of 20–25 percent of responses from previous assessments (for identical items) is systematically interspersed among current responses for rescoring. If necessary, current assessment results are adjusted to account for any differences.

To test scoring reliability, constructed-response item score statistics are calculated for the portion of responses that are scored twice. Cohen's Kappa is the reliability estimate used for dichotomized items and the intraclass correlation coefficient is used as the index of reliability for nondichotomized items. Scores are also constructed for items that are rescored in a later assessment. For example, some 2007 reading and mathematics items were rescored in 2009.

Editing. The first phase of data editing takes place during the keying or scanning of the survey instruments. Machine edits verify that each sheet of each document is present and that each field has an appropriate value. The edit program checks each booklet number against the session code for appropriate session type, the school code against the control system record, and other data fields on the booklet cover for valid ranges of values. It then checks each block of the document for validity, proceeding through the items within the block. Each piece of input data is checked to verify that it is of an acceptable type, that the value falls within a specified range of values, and that it is consistent with other data values. At the end of this process, a paper edit listing of data errors is generated for nonimage and key-entered documents. Image-scanned items requiring correction are displayed at an online editing terminal.

In the second phase of data editing, experienced editing staff review the errors detected in the first phase, compare the processed data with the original source document, and indicate whether the error is correctable or noncorrectable per the editing specifications. Suspect items found to be correct as stated, but outside the edit specifications, are passed through modified edit programs. For nonimage and key-entered documents, corrections are made later via key-entry. For image-processed documents, suspect items are edited online. The edit criteria for each item in question appear on the screen along with the item, and corrections are made immediately. Two different people view the same suspect item and operate on it separately; a "verifier" ensures that the two responses are the same before the system accepts that item as correct.

For assessment items that must be paper-scored rather than scored using the image system (as was the case for some mathematics items in the 1996 NAEP), the score sheets are scanned on a paper-based scanning system and then edited against tables to ensure that all responses were scored with only one valid score and that only raters qualified to score an item were allowed to score it. Any discrepancies are flagged and resolved before the data from that scoring sheet are accepted into the scoring system.

In addition, a count-verification phase systematically compares booklet IDs with those listed in the NAEP administration schedule to ensure that all booklets expected to be processed were actually processed. Once all corrections are entered and verified, the corrected records are pulled into a mainframe data set and then re-edited with all other records. The editing process is repeated until all data are correct.

Estimation Methods

Once NAEP data are scored and compiled, the responses are weighted according to the sample design and population structure and then adjusted for nonresponse. This ensures that students' representation in NAEP matches their actual proportion of the school population in the grades assessed. The analyses of NAEP data for most subjects are conducted in two phases: scaling and estimation. During the scaling phase, item response theory (IRT) procedures are used to estimate the measurement characteristics of each assessment question. During the estimation phase, the results of the scaling are used to produce estimates of student achievement (proficiency) in the various subject areas. Marginal maximum likelihood (MML) methodology is then used to estimate characteristics of the proficiency distributions. Estimates of student achievement are included in the NAEP database; estimates of other variables are not included.

Weighting. The weighting for the national and state samples reflects the probability of selection for each student in the sample, adjusted for school and student nonresponse. The weight assigned to a student's responses is the inverse of the probability that the student would be selected for the sample. Prior to 2002, poststratification was used to ensure that the weighting was representative of certain subpopulations corresponding to figures from the U.S. Census and the Current Population Survey (CPS).

Student base weights. The base weight assigned to a student is the reciprocal of the probability that the student would be selected for a particular assessment. This probability is the product of the following two factors:

- the conditional probability that the school would be selected, given the strata; and

- the conditional probability, given the school, that the student would be selected within the school.

Nonresponse adjustments of base weights. The base weight for a selected student is adjusted by two nonresponse factors. The first factor adjusts for sessions that were not conducted. This factor is computed separately within classes formed by the first three digits of strata (formed by crossing the major stratum and the first socioeconomic characteristic used to define the final stratum). Occasionally, additional collapsing of classes is necessary to improve the stability of the adjustment factors, especially for the smaller assessment components. The second factor adjusts for students who failed to appear in the scheduled session or makeup session. This nonresponse adjustment is completed separately for each assessment. For assessed students in the trend samples, the adjustment is made separately for classes of students based on subuniverse and modal grade status. For assessed students in the main samples, the adjustment classes are based on subuniverse, modal grade status, and race class. In some cases, nonresponse classes are collapsed into one class to improve the stability of the adjustment factors.

NIES Part II weighting. In the NIES Part II,. the school probability of selection is a function of three factors: NAEP selection, the probability of being retained for NIES Part II, and the number of AI/AN students in the NAEP sample per school. Nonresponse adjustments at the school level attempt to mitigate the impact of differential response by school type (public, private, and BIE), region, and estimated percentage enrollment of AI/AN students. For student weights, nonresponse adjustments take into account differential response rates based on student age (above age for grade level or not) and English language learner status. In order to partially counteract the negative impact of low private school participation, a poststratification adjustment is applied to the NIES Part II weights. The relative weighted proportions of students from public, private, and BIE schools, respectively, are adjusted to match those from the NIES Part I data. This not only ensured greater consistency between the findings of the two NIES components, but since the proportions of students are more reliably estimated from the NIES Part I data (which involved a far larger school sample

than Part II), this weight adjustment increases the accuracy and reliability of the NIES Part II results.

Scaling. For purposes of summarizing item responses, ETS developed a scaling technique that has its roots in IRT procedures and the theories of imputation of missing data.

The first step in scaling is to determine the percentage of students who give various responses to each cognitive, or subject-matter, question and each background question. For cognitive questions, a distinction is made between missing responses at the end of a block (i.e., missing responses after the last question the student answered) and missing responses before the last observed response. Missing responses before the last observed response are considered intentional omissions. Missing responses at the end of a block are generally considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented that question are used in the analysis. Each cognitive question is also examined for differential item functioning (DIF). DIF analyses identify questions on which the scores of different subgroups of students at the same ability level differ significantly.

Development of scales. Separate subscales are derived for each subject area. For the main assessments, the frameworks for the different subject areas dictate the number of subscales required. In the 2009 NAEP, five subscales were created for the main assessment in mathematics in grades 4 and 8 (one for each mathematics content strand), and three subscales were created for science (one for each field of science: Earth, physical, and life). A composite scale is also created as an overall measure of students' performance in the subject area being assessed (e.g., mathematics). The composite scale is a weighted average of the separate subscales for the defined subfields or content strands. For the long-term trend assessments, a separate scale is used for summarizing proficiencies at each age in mathematics and reading.

Within-grade vs. cross-grade scaling. The reading and mathematics main NAEP assessments were developed with a cross-grade framework, where the trait being measured was conceptualized as cumulative across the grades of the assessment. Accordingly, a single 0−500 scale was established for all three grades in each assessment. In 1993, however, NAGB determined that future NAEP assessments should be developed using within-grade frameworks and be scaled accordingly. This both removed the constraint that the trait being measured is cumulative and eliminated the need for overlap of questions across grades. Any questions that happen to be the same across grades are scaled separately for each grade, thus making it possible for common questions to function differently in the separate grades.

The 1994 history and geography assessments were developed and scaled within grade, according to NAGB's new policy. The scales were aligned so that grade 8 had a higher mean than grade 4 and grade 12 had a higher mean than grade 8. The 1994 reading assessment, however, retained a cross-grade framework and scaling. All three main assessments in 1994 used scales ranging from 0 to 500.

The 2008 long-term trend assessments remained cross-age, using a 0−500 scale. The 2009 main science assessment was developed within-grade, but adopted new scales ranging from 0 to 300. The 2005 main assessment in mathematics continued to use a cross-grade framework with a 0−500 scale in grades 4 and 8, but used a 0-300 within-grade scale. In 1998, reading, writing and civics assessments were scaled within-grade.

Linking of scales. Before 2002, results for the main state assessments were linked to the scales for the main national assessments, enabling state and national trends to be studied. Equating the results of the state and national assessments depended on those parts of the main national and state samples that represented a common population: (1) the state comparison sample—students tested in the national assessment who come from the jurisdictions participating in the state NAEP; and (2) the state aggregate sample—the aggregate of all students tested in the state NAEP. Since 2002, the national sample has been a superset of the state samples (except in those states that do not participate). Thus, equating is not necessary.

Imputation. **U**ntil the 2002 NAEP assessment, no statistical imputations were generated for missing values in the teacher, school, or SD/ELL questionnaires, or for missing answers to cognitive questions. Most answers to cognitive questions are missing by design. For example, 8^{th}-grade students being assessed in reading are presented with, on average, 21 of the 110 assessment items. Whether any given student gets any of the remaining 89 individual questions right or wrong is not something that NAEP imputes. However, since 1984, multiple imputation

techniques have been used to create plausible values. Once created, subsequent users can analyze these plausible values with common software packages to obtain NAEP results that properly account for NAEP's complex item sampling designs.

Because no student takes even a quarter of the questions in an assessment, individual scores cannot be calculated. Trying to use partial scores based on the small proportion of the assessment to which any given student is exposed would lead to biased results for groups scores due to an inherently large component of measurement error. NAEP developed a process of group score calculation in order to get around the unreliability and noncomparability of NAEP's partial test forms for individuals. NAEP estimates group score distributions using MML estimation, a method that calculates group score distributions based directly on each student's responses to cognitive questions, not on summary scores for each student. As a result, the unreliability of individual-level scores does not decrease NAEP's accuracy in reporting group scores. The MML method does not employ imputations of answers to any questions or of scores for individuals.

Imputation is performed in three stages. The first stage requires estimating IRT parameters for each cognitive question. The second stage results in MML estimation of a set of regression coefficients that capture the relationship between group score distributions and nearly all the information from the variables in the teacher, school, or SD/ELL questionnaires, as well as geographical, sample frame, and school record information. The third stage involves the imputation that is designed to reproduce the group-level results that could be obtained during the second stage.

NAEP's imputations follow Rubin's (1987) proposal that the imputation process be carried out several times, so that the variability associated with group score distributions can be accurately represented. NAEP estimates five plausible values for each student. The five plausible values are calculated using the regression coefficients estimated in the second stage. Each plausible value is a random selection from the joint distribution of potential scale scores that fit the observed set of response for each student and the scores for each of the groups to which each student belongs. Estimates based on plausible values are more accurate than if a single (necessarily partial) score were to be estimated for each student and averaged to obtain estimates of subgroup performances. Using the plausible values eliminates the need for secondary analysts to have access to specialized MML software and ensures that the estimates of average performance of groups and estimates of variability in those averages are accurate.

Recent Changes

Several important changes have been implemented since 1990.

- Beginning with the 1990 mathematics assessment, NAGB established three reporting levels for reporting NAEP results: basic, proficient, and advanced.
- In 1990, state assessments were added to NAEP. The 1990 to 1994 assessments are referred to as trial state assessments.
- In 1992, a generalized partial-credit model (GPCM) was introduced to develop scales for the more complex constructed-response questions. The GPCM model permits the scaling of questions scored according to multipoint rating schemes.
- In 1993, NAGB determined that future NAEP assessments should have within-grade frameworks and scales. The 1994 main history and geography assessments followed this new policy, as did the 1996 main science assessment, and the 1998 writing assessment. Mathematics and reading in the main NAEP will continue to have cross-grade scales until further action by NAGB (and a parallel change in the trend assessment), except for mathematics at grade 12, which was removed from cross-grade scales and reported in a within-grade scale in 2005.
- In 1994, the new image-based scoring system virtually eliminated paper handling during the scoring process. This system also permits scoring reliability to be monitored online and recalibration methods to be introduced.
- The 1996 main NAEP included new samples for the purpose of studying greater inclusion of SD/LEP students and obtaining data on students eligible for advanced mathematics or science sessions.
- In 1997, there was a probe of student performance in the arts.
- New assessment techniques included: open-ended items in the 1990 mathematics assessment; primary trait, holistic, and writing mechanics scoring procedures in the 1992 writing assessment; the use of calculators in the 1990, 1992, 1996, and 2000 mathematics assessments;

a special study on group problem solving in the 1994 history assessment; and a special study in theme blocks in the 1996 mathematics and science assessments.

- Beginning in 1998, testing accommodations were provided in the NAEP reading assessments; in this transition to a more inclusive NAEP, administration procedures were introduced that allowed the use of accommodations (e.g., extra time, individual rather than group administration) for students who required them to participate. During this transition period, reading results in 1998 were reported for two separate samples: one in which accommodations were not permitted and one in which accommodations were permitted. Beginning in 2002, accommodations were permitted for all reading administrations.

- In 1999, NAGB discontinued the long-term trend assessment in writing for technical reasons. More recently, NAGB decided that changes were needed to the design of the science assessment and, given recent advances in the field of science, to its content. As a result, the science long-term trend assessment was not administered in 2003-04.

- With the expansion and redesign of NAEP under the No Child Left Behind Act, NAEP's biennial state-level assessments are being administered by contractor staff (not local teachers). The newly redesigned NAEP has four important features. First, NAEP is administering tests for different subjects (such as mathematics, science, and reading) in the same classroom, thereby simplifying and speeding up sampling, administration, and weighting. Second, NAEP is conducting pilot tests of candidate items for the next assessment and field tests of items for precalibration in advance of data collection, thereby speeding up the scaling process. Third, NAEP is conducting bridge studies, administering tests both under the new and the old conditions, thereby providing the possibility of linking old and new findings. Finally, NAEP is adding additional test questions at the upper and lower ends of the difficulty spectrum, thereby increasing NAEP's power to measure performance gaps.

- Beginning in 2002, the NAEP national sample for main national assessment was obtained by aggregating the samples from each state, rather than by obtaining an independently selected national sample. Prior to 2002, separate samples were drawn for the NAEP main national and state assessments.

- In 2002, TUDA began assessing performance in five large urban districts with reading and writing assessments. TUDA continued in 2003 in nine large urban districts with reading and mathematics and in 2005 in 10 large urban districts with reading, mathematics, and science.

- Beginning with the 2003 NAEP, each state must have participation from at least 85 percent—instead of 70 percent—of the schools in the original sample in order to have its results reported.

- In 2003 and 2005, Puerto Rico participated in the NAEP assessment of mathematics. However, Puerto Rico was excused from the NAEP assessment of reading in English because Spanish is the language of instruction in Puerto Rico. NCES also administered the 2007 mathematics assessment in Puerto Rico.

- In 2004, several changes were implemented to the NAEP long-term trend assessments to reflect changes in NAEP policy, maintain the integrity of the assessments, and increase the validity of the results obtained. The changes to the assessment instruments include: removal of science items; inclusion of students with disabilities and English language learners; replacement of items that used outdated contexts; creation of a separate background questionnaire; elimination of "I don't know" as a response option for multiple-choice items; and use of assessment booklets that pertain to a single subject area (whereas in the past, a single assessment booklet may have contained both reading and mathematics items).

- In 2005, NAGB introduced changes in the NAEP mathematics framework for grade 12 in both the assessment content and administration procedures. One of the major differences between the 2005 assessment and previous assessments at grade 12 is the five content areas were collapsed into four areas, with geometry and measurement being combined. In addition, the assessment included more questions on algebra, data analysis, and probability to reflect changes in high school mathematics standards and coursework. The overall average mathematics score in 2005 was set at 150 on a 0–300 scale.

- In 2006, economics was assessed at grade 12 for the first time. A within-grade scale was developed, with the overall average economics score in 2006 set at 150 on a 0–300 scale.

- In 2009, TUDA was expanded to 18 large urban districts, assessing reading, mathematics and science. In addition, 11 states were assessed in reading and mathematics at grade 12 on a trial basis.

- In 2009, interactive computer tasks in science were administered online at grades 4, 8, and 12. These tasks consisted of simulations for the students to draw inferences and conclusions about a problem.

Future Plans

The next trend assessment will be administered in 2012, and then every 4 years thereafter. Main assessments are scheduled for annual administration. Reading and mathematics are assessed every 2 years in odd-numbered years; science and writing are scheduled to be assessed every 4 years (in the same years as reading and mathematics, but alternating with each other); and other subjects are assessed at the national level in even-numbered years. Writing will be assessed online in 2011 to a national sample of 4th and 8th graders.

5. DATA QUALITY AND COMPARABILITY

As the Nation's Report Card, NAEP must report accurate results for populations of students and subgroups of these populations (e.g., Black, Hispanic, or other race/ethnicity, or students attending nonpublic schools). Although only a very small percentage of the student population in each grade is assessed, NAEP estimates are accurate because they depend on the absolute number of students participating, not on the relative proportion of students.

Every activity in NAEP assessments is conducted with rigorous quality control, contributing both to the quality and comparability of the assessments and their results. All questions undergo extensive reviews by subject-area and measurement specialists, as well as careful scrutiny to eliminate any potential bias or lack of sensitivity to particular groups. The complex process by which NAEP data are collected and processed is monitored closely. Although each participating state is responsible for its own data collection for the main state NAEP, Westat ensures uniformity of procedures across states through training, supervision, and quality control monitoring.

With any survey, however, there is the possibility of error. The most likely sources of error in NAEP are described below.

Sampling Error

Two components of uncertainty in NAEP assessments are accounted for in the variability of statistics based on scale scores: (1) the uncertainty due to sampling only a small number of students relative to the whole population; and (2) the uncertainty due to sampling only a relatively small number of questions. The variability of estimates of percentages of students having certain back-ground characteristics or answering a certain cognitive question correctly is accounted for by the first component alone.

Because NAEP uses complex sampling procedures, a jackknife replication procedure is used to estimate standard errors. While the jackknife standard error provides a reasonable measure of uncertainty about student data that can be observed without error, each student in NAEP assessments typically responds to so few questions within any content area that the scale score for the student would be imprecise. It is possible to describe the performance of groups and subgroups of students because, as a group, all students are administered a wide range of items.

NAEP uses MML procedures to estimate group distributions of scores. However, the underlying imprecision that makes this step necessary adds an additional component of variability to statistics based on NAEP scale scores. This imprecision is measured by the imputed variance, which is estimated by the variance among the plausible values drawn from each student's posterior distribution of possible scores. The final estimate of the variance is the sum of the sampling variance and the measurement variance.

Nonsampling Error

While there is the possibility of some coverage error in NAEP, the two most likely types of nonsampling error are nonresponse error due to nonparticipation and measurement error due to instrumentation defects (described below). The overall extent of nonsampling error is largely unknown.

Coverage error. In NAEP, coverage error can result either from the sampling frame of schools being incomplete or from the schools' failure to include all the students on the lists from which grade or age samples are drawn. For the 2009 NAEP, the 2008

school list maintained by CCD supplied the names of the regular public schools, BIE schools, and DoDEA schools. This list, however, did not include schools that opened between 2008 and the time of the 2009 NAEP. To be sure that students in new public schools were represented, each sample district in NAEP was asked to update lists of schools with newly eligible schools.

Catholic and other nonpublic schools in the 2009 NAEP were obtained from the PSS. PSS uses a dual-frame approach. The list frame (containing most private schools in the country) is supplemented by an area frame (containing additional schools identified during a search of randomly selected geographic areas around the country). Coverage of private schools in the PSS is very high. (See chapter 3.)

Nonresponse error.

Unit nonresponse. In the 2009 reading and mathematics assessments, all 52 states and jurisdictions[3] met participation rate standards at both grade 4 and grade 8. The national school participation rates for public and private schools combined were 97 percent at grades and grade 8. Student participation rates were 95 percent at grade 4 and 93 percent at grade 8. Participation rates needed to be 70 percent or higher to report results separately for private schools. While the participation rate for private schools did meet the standard in 2009, it did not always meet the standard in previous assessment years. See table 11 for more details.

In the 2007 reading and mathematics assessments, all 52 states and jurisdictions[4] met participation rate standards at both grades 4 and 8. The national school participation rates for public and private schools combined were 98 percent at grade 4 and 97 percent at grade 8. Student participation rates were 95 percent at grade 4 and 92 percent at grade 8. Participation rates needed to be 70 percent or higher to report results separately for private schools.

In the 2005 reading and mathematics assessments at grade 12, participation standards were met for public schools but not for private schools. At the student level, response rates at grade 12 fell below 85 percent for students in both public and private schools. A nonresponse bias analysis showed significant differences between responding and nonresponding public school students in terms of gender, race/ethnicity, age, and English language learner identification. Although the differences are quite small, it is unlikely that nonresponse weighting adjustments completely accounted for these differences.

In the 2008 trend assessments, private school participation rate at age 17 was 61 percent, below the standard for reporting. However, Catholic school participation rates at all three ages (88, 94, and 76 percent at ages 9, 13, and 17, respectively) met the reporting standards.

In the 2007 NIES Part II, questionnaires were completed by about 10,400 grade 4 students from 1,700 schools and 11,300 grade 8 students from 1,800 schools. Also responding to the survey were about 3,000 grade 4 teachers, 4,600 grade 8 teachers, 1,700 grade 4 school administrators and 1,800 grade 8 school administrators associated with these students. Some school administrators responded for both grades 4 and 8. The weighted student response rates were 85 percent at grade 4 and 82 percent at grade 8. The weighted school response rates were 88 percent at grade 4 and 90 percent at grade 8.

In the 2005 NIES Part II, questionnaires were completed by about 2,600 grade 4 students and 2,500 grade 8 students at approximately 480 schools. Also responding to the survey were about 480 grade 4 teachers, 820 grade 8 teachers, 240 grade 4 principals, and 230 grade 8 principals associated with these students. Some principals responded for both grades 4 and 8. The weighted student response rates were 95 percent at grade 4 and 91 percent at grade 8. The weighted school response rates were 87 percent at grade 4 and 93 percent at grade 8.

In the 2004 long-term trend reading and mathematics assessments, the overall response rate (the product of the weighted school participation rate before substitution and the weighted student participation rate) fell below the NCES reporting target of 85 percent for ages 13 and 17 at the school level and for age 17 at the student level. At age 13, a bias was found for private schools, as a greater proportion of nonresponses were from other private schools than from Catholic schools. In addition, nonrespondent schools in the long-term trend assessment had a lower percentage of Black students than participating schools. Likewise, at age 17, private schools were disproportionately less likely to participate, and within private schools, Catholics and Conservative Christian schools had higher participation rates than other private schools. Nonrespondent schools

[3] It includes 50 states, District of Columbia, and DoDEA.
[4] It includes 50 states, District of Columbia, and DoDEA.

Table 11. Weighted school, student, and overall response rates for selected NAEP national assessments, by assessment and grade: 2006-2009

Assessment and grade	School participation[1]		Student participation	
	Student weighted	School weighted	Student weighted	Overall participation
2009 Mathematics				
Grade 4	97	91	95	92
Grade 8	97	87	93	90
2009 Reading				
Grade 4	97	91	95	92
Grade 8	97	87	93	90
2009 Science				
Grade 4	97	91	95	92
Grade 8	97	87	93	90
Grade 12	83	79	80	66
2008 Trend				
Age 9	96	91	95	91
Age 13	95	89	94	89
Age 17	90	85	88	79
2007 Writing				
Grade 8	97	87	92	90
Grade 12	89	83	80	71
2007 Reading				
Grade 4	98	92	95	93
Grade 8	97	87	92	90
2007 Mathematics				
Grade 4	98	92	95	93
Grade 8	97	87	92	90
2006 Economics				
Grade 12	79	78	73	58
2006 Civics				
Grade 4	92	86	95	88
Grade 8	93	86	92	85
Grade 12	79	78	72	57
2006 U.S. history				
Grade 4	91	88	95	87
Grade 8	91	85	92	84
Grade 12	80	80	73	59

[1] Participation rates do not include substitutions.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Mathematics, Reading and Science Assessments, 2008 Trend Assessment, 2007 Writing, Reading and Mathematics Assessments, 2006 Economics, Civics, and U.S. history Assessments.

also had a slightly higher percentage of Asian students than participating schools at age 17. At the student level at age 17, some bias was shown for race/ethnicity, free lunch eligibility, and disability status.

Item nonresponse. Specific information about nonresponse for particular items is available in NAEP summary data tables on the Web http://nces.ed.gov/nationsreportcard/naepdata/.

Measurement error. Nonsampling error can result from the failure of the test instruments to measure what is being taught and, in turn, what is being learned by students. For example, the instruments may contain ambiguous definitions and/or questions that lead to different interpretations by students. Additional sources of measurement error are the inability or unwillingness of students to give correct information and errors in the recording, coding, or scoring of data.

To assess the quality of the data in the final NAEP database, survey instruments are selected at random and compared, character by character, with their records in the final database. As in past years, the 2000 NAEP data-base was found to be more than accurate enough to support analyses.

The observed error rates for the 2000 NAEP were comparable to those of past assessments. Error rates ranged from 8 errors per 10,000 responses for the Teacher Survey questionnaire to 44 errors per 10,000 responses for the School Characteristics and Policies Survey questionnaire.

Revised results. Following the 1994 NAEP assessment, two technical problems were discovered in the procedures used to develop the scale and achievement levels for the 1990 and 1992 mathematics assessments. These errors affected the mathematics scale scores reported for 1992 and the achievement-level results reported for 1990 and 1992.

NCES and NAGB evaluated the impact of these errors and subsequently reanalyzed data and reported the revised results from both mathematics assessments. The revised results for 1990 and 1992 are presented in the 1996 mathematics reports. For more detail on these problems, see *The NAEP 1996 Technical Report* (Allen, Carlson, and Zelenak 1999) and the *Technical Report of the NAEP 1996 State Assessment Program in Mathematics* (Allen et al. 1997).

There were also problems related to reading scale scores and achievement levels. These errors affected the 1992 and 1994 NAEP reading assessment results. The 1992 and 1994 reading data have been reanalyzed and reissued in revised reports. For more information, refer to *The NAEP 1994 Technical Report* (Allen, Kline, and Zelenak 1996) and the *Technical Report of the NAEP 1994 Trial State Assessment in Reading* (Mazzeo, Allen, and Kline 1995).

Data Comparability

NAEP allows reliable comparisons between state and national data for any given assessment year. By linking scales across assessments, it is possible to examine short-term trends for data from the main national and state NAEP and long-term trends for data from the long-term trend NAEP.

Main national vs. main state comparisons. NAEP data are collected using a closely monitored and standardized process, which helps ensure the comparability of the results generated from the main national and state assessments. The main national NAEP and main state NAEP use the same assessment booklets, and, since 2002, they have been administered in the same sessions using identical procedures.

Short-term trends. Although the test instruments for the main national assessments are designed to be flexible and thus adaptable to changes in curricular and educational approaches, they are kept stable for shorter periods (up to 12 years or more) to allow analysis of short-term trends. For example, through common questions, the 1996 main national assessment in mathematics was linked to both the 1992 and 1994 assessments.

For 2005, NAGB adopted a new mathematics framework for grade 12 to reflect changes in high school standards and coursework. In addition, changes were made in booklet design and calculator-use policy for the one-third of the assessment in which calculators were allowed. As a result of these changes, the 2005 results could not be placed on the previous NAEP scale and are not compared to results from previous years.

Long-term trends. In order to make long-term comparisons, the long-term trend NAEP uses different samples than the main national NAEP. Unlike the test instruments for the main NAEP, the long-term instruments in mathematics and reading have remained relatively unchanged from those used in previous assessments. The 2004 trend instruments were almost identical to those used in the 1970s. The trend NAEP allows the measurement of educational

progress since 1971 in reading and 1973 in math ematics. For more detail on the linking of scales in the trend NAEP, see “Scaling” in section 4 above.

The long-term trend assessment was updated in several ways in 2004 (e.g., inclusion of SD/ELL students). To ensure the comparability of the new assessment and the previous assessments, a bridge study was performed.

Linking to non-NAEP assessments. Linking results from the main state assessments to those from the main national assessments has encouraged efforts to link NAEP assessments with non-NAEP assessments.

Linking to state assessment. NAEP data can be used to map state proficiency standards in reading and mathematics onto the appropriate NAEP scale. The mapping exercise was carried out for data from the 2004–05 and 2006-07 academic years at both grades 4 and 8. For each of the four subject and grade combinations, the NAEP score equivalents to the states' proficiency standards vary widely, spanning a range of 60 to 80 NAEP score points. Although there is an essential ambiguity in any attempt to place state standards on a common scale, the ranking of the NAEP score equivalents to the states' proficiency standards offers an indicator of the relative stringency of those standards. There are plans to do this mapping for the 2008-09 school year also.

There is a strong negative correlation between the proportions of students meeting the states' proficiency standards and the NAEP score equivalents to those standards, suggesting that the observed heterogeneity in states' reported percents proficient can be largely attributed to differences in the stringency of their standards. There is, at best, a weak relationship between the NAEP score equivalents to the states' proficiency standards and the states' average scores on NAEP. Finally, most of the NAEP score equivalents fall below the cut-point corresponding to the NAEP proficient level, and many fall below the cut-point corresponding to the NAEP basic level.

These results should be employed cautiously, as differences among states in apparent stringency can be due, in part, to reasonable differences in the assessment frameworks, the types of item formats employed, and the psychometric characteristics of the tests. Moreover, there is some variation among states in the proportion of NAEP sample schools that could be employed in the analysis.

Linking to the International Assessment of Educational Progress (IAEP). In 1992, results from the 1992 NAEP assessment in mathematics in grade 8 were successfully linked to those from IAEP of 1991. Sample data were collected from U.S. students who had been administered both instruments. The relation between mathematics proficiency in the two assessments was modeled using regression analysis. This model was then used as the basis for projecting IAEP scores from non-U.S. countries onto the NAEP scale. *The relation between the IAEP and NAEP assessments was relatively strong and could be modeled well. The results, however, should be considered only in the context of the similar construction and scoring of the two assessments. Further studies should be initiated cautiously, even though the path to linking assessments is now better understood.*

Linking to TIMSS. The success in linking NAEP to the IAEP sparked an interest in linking the results from the 1996 NAEP assessments in mathematics and science in grade 8 to those from the Third International Mathematics and Science Study (TIMSS) of 1995. The data from this study became available at approximately the same time as the1996 NAEP data for mathematics and science. Because the two assessments were conducted in different years and no students responded to both assessments, the regression procedure that linked NAEP and IAEP assessments could not be used. The results from grade 8 NAEP and TIMSS assessments were instead linked by matching their score distributions. A comparison of the linked results with actual results from states that participated in both assessments suggested that the link was working acceptably. *The results from U.S. students were linked to those of their academic peers in more than 40 other countries. As with the IAEP linked results, these results should be used cautiously.*

A second study attempted to link the 2000 grade 8 NAEP assessments in mathematics and science to the 1999 grade 8 TIMSS (which also assessed mathematics and science). The primary linkage used a projection method, which drew data from a sample of students to whom both assessments were administered. The linkage found that the projections were substantially off the mark. A secondary linkage, based on nationally reported numbers using a statistical moderation approach, provided a fairly weak linkage; the moderation linkage did a decent job of projecting TIMSS scores from NAEP scores in the 12 states that participated in both studies, but failed to predict the TIMSS score in the linking sample.

The analyses showed that the TIMSS assessments functioned differently in the linking sample than they did in the national and state samples. A recent study (Phillip 2009) shows that it is possible to make comparisons between TIMSS 2007 and NAEP 2007. For more details, please refer to *The Second Derivative: International Benchmarks in Mathematics for U.S. States and School Districts* (Phillip 2009).

Comparisons with TIMSS. Studies were undertaken to compare the content of two fourth- and eighth-grade assessments in mathematics and science: the NAEP 2000 assessment and the TIMSS 2003 assessment. The comparison study drew upon information provided by the developers of the assessments, as well as data obtained from an expert panel convened to compare the frameworks and items from the two assessments on various dimensions.

For science, the content comparisons between NAEP and TIMSS reveal some key differences in the topics covered, grade-level correspondence, and the characteristics of the item pools on other dimensions. All of these factors together may result in differences in student performance, and it is important to consider these differences when interpreting the results from the different assessments.

Differences in the science content included in each assessment can be seen at both the framework level and in the pool of items developed based on these frameworks. Even in content areas where there is considerable overlap of the frameworks (such as life science and Earth science), a closer examination of the topics and specific objectives covered by the items in each assessment reveals some important differences. In comparison to NAEP, whose framework was developed in the context of the U.S. system, the TIMSS framework reflects a consensus across many countries. Some of the differences in curricula across these countries are reflected in the frameworks and in the differences in content of the two assessments. In particular, the inclusion in TIMSS of separate content areas in chemistry, physics, and environmental science results in broader topic coverage in some areas. While there is a considerable overlap in the topics included in some content areas, the items included in each assessment place different emphases at the topic level. In addition, the “hands-on” tasks in NAEP provide complementary information to the pencil-and-paper portions of both assessments, enabling the measurement of student performance in this area of knowing and doing science.

With respect to mathematics, a comparison of the frameworks revealed considerable agreement on the general boundaries and basic organization of mathematics content, with both assessments including five main content areas corresponding to traditional mathematics curricular areas: number, measurement, geometry, data, and algebra. Both the NAEP and TIMSS frameworks also include dimensions that define a range of cognitive skills and processes that overlap the two assessments. Despite these apparent similarities at the broadest level, a closer examination of the items in each assessment reveals different emphases at the topic and subtopic levels, as well as some differences in grade-level expectations across mathematics topics.

Comparisons with PIRLS. In 2003, NCES released results for both the 2001 Progress in International Reading Literacy Study (PIRLS) fourth-grade assessment and the 2002 NAEP fourth-grade reading assessment. In anticipation of questions about how these two assessments compare, NCES convened an expert panel to compare the content of the PIRLS and NAEP assessments and determine if they are measuring the same construct. This involved a close examination of how PIRLS and NAEP define reading, the texts used as the basis for the assessments, and the reading processes required of students in each. The comparison suggests that there is a great deal of overlap in what the two assessments are measuring. While they do seem to define and measure the same kind of reading, PIRLS is an easier assessment than NAEP, with more text-based tasks and shorter, less complex reading passages. The similarities and differences between the two are discussed below.

The comparison revealed that, overall, the NAEP and PIRLS reading assessments are quite similar. Both define reading similarly, as a constructive process. Both use high-quality reading passages and address similar purposes for which young children read (for literary experience and information). Both call for students to develop interpretations, make connections across text, and evaluate aspects of what they have read. Finally, both have a similar distribution of multiple-choice and constructed-response items: in each, about half of the items are constructed-response items.

While the two assessments have similar definitions of reading and assess many of the same aspects of it, a closer look at how the domain is operationalized by each revealed some important differences. NAEP places more emphasis than PIRLS on having students taking what they have read and connecting it to other readings or knowledge. PIRLS places a greater

emphasis than NAEP on text-based reading skills and interactions, including items that ask students to locate information in the text, make text-based inferences and interpretations, and evaluate aspects of the text.

The PIRLS reading passages are, on average, about half the length of the NAEP reading passages. PIRLS readability formulas indicate that the passages used in PIRLS are less complex than those used in NAEP. The classification of items also revealed differences in how the two frameworks function. The panel had an easier time classifying PIRLS and NAEP items by the PIRLS framework categories than by the NAEP framework categories. For more information on the similarities and differences between PIRLS and NAEP, see *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments* (Binkley and Kelly 2003).

Comparisons with the International Association for the Evaluation of Educational Achievement's (IEA) Reading Literacy Study. The picture of American students' reading proficiency provided by NAEP assessments is less optimistic than that indicated by the IEA Reading Literacy Study. This can be explained by the following:

(1) *The basis for reporting differs considerably between the two assessments.* With the IEA study, students are compared against other students and not against a standard set of criteria on knowledge, as in NAEP. Much of NAEP reporting is based on comparisons between actual student performance and desired performance (what students are expected to do).

(2) *NAEP and IEA assess different aspects of reading.* More than 90 percent of the IEA items assess tasks covered in only 17 percent of NAEP items. Furthermore, virtually all of the IEA items are aimed solely at literal comprehension and interpretation, while such items make up only one-third of NAEP reading assessments.

(3) *NAEP and IEA differ in what students must do to demonstrate their comprehension.* More interpretive and higher level thinking is required to reach the advanced level in NAEP than in the IEA study. Also, NAEP requires students to generate answers in their own words much more frequently than does the IEA study. Moreover, the IEA test items do not cover the entire expected ability range. Many American students answer every IEA item correctly, making it impossible to distinguish between the abilities of students in the upper range. In contrast, the range of item difficulty on NAEP reading assessments exceeds the ability of most American students, so differences in the abilities of students in the upper range can be distinguished easily.

Despite the differences between these two assessments, there is a high probability that, if students from other countries were to take NAEP, the rank ordering or relative performance of countries would be about the same as in the IEA findings. This assumption is based on the theoretic underpinnings of item response theory and its application to the test scaling used for both the IEA Reading Literacy Study and the NAEP reading assessment.

6. CONTACT INFORMATION

For content information on NAEP, contact:

Peggy Carr
Phone: (202) 502-7321
E-mail: peggy.carr@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Allen, N.L., Carlson, J.E., and Zelenak, C.A. (1999). *The NAEP 1996 Technical Report* (NCES 1999-452). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Allen, N.L., Donoghue, J.R., and Schoeps, T.L. (2001). *The NAEP 1998 Technical Report* (NCES 2001-509). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Allen, N.L., Jenkins, F., Kulick, E., and Zelenak, C. A. (1997). *Technical Report of the NAEP 1996 State Assessment Program in Mathematics* (NCES 97-

951). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Allen, N.L., Kline, D.L., and Zelenak, C.A. (1996). *The NAEP 1994 Technical Report* (NCES 97-897). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Allen, N.L., Swinton, S.S., Isham, S.P., and Zelenak, C.A. (1998). *Technical Report: NAEP 1996 State Assessment Program in Science* (NCES 98-480). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Alt, M.N., and Bradby, D. (1999). *Procedures Guide for Transcript Studies* (NCES Working Paper 1999-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Braun, H., Jenkins, F., and Grigg, W. (2006). *Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling* (NCES 2006-461). U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. Washington, DC: U.S. Government Printing Office.

Calderone, J., King, L.M., and Horkay, N. (1997). *The NAEP Guide: A Description of the Content and Methods of the 1997 and 1998 Assessments* (NCES 97-990). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Johnson, C. (2004). *Nation's Report Card: An Overview of NAEP* (NCES 2004-552). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Jonhson, E.G., and Carlson, J.E. (1994). *NAEP 1992 Technical Report* (NCES 94-490). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jonhson, E.G., Mazzeo, J., and Kline, D.L. (1994). *Technical Report of the NAEP 1992 Trial State Assessment in Reading* (NCES 94-472). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Lutkus, A.D., Mazzeo, J., Zhang, J., Jerry, L., Barton, K.E., and Zenisky, A. (2003). *Including Special-Needs Students in the NAEP 1998 Reading Assessment, Part I, Comparison of Overall Results With and Without Accommodations* (NCES 2003-467). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Mazzeo, J., Allen, N.L., and Kline, D.L. (1995). *Technical Report of the NAEP 1994 Trial State Assessment in Reading* (NCES 96-116). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Moran, R., and Rampey, B. (2008). *National Indian Education Study 2007 - Part II: The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8* (NCES 2008-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Patricia, D. (2005). *The Nation's Report Card: An Introduction to The National Assessment of Educational Progress (NAEP)* (NCES 2005-454). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Perie, M., Vanneman, A., and Goldstein, A. (2005). *Student Achievement in Private Schools: Results From NAEP 2000–2005* (NCES 2006-459). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

Persky, H., Brent, S.A., and Janice, A. (2003). *Assessing the Arts: Selected NAEP Tasks and Scoring Guides for Grades 4 and 12 1997 Field Test. Dance, Music, Theatre, and Visual Art* (NCES 2003-452). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Phillips, G., and Goldstein, A. (1996). *Technical Issues in Large-Scale Performance Assessment* (NCES 96-802). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Roey, S., Caldwell, N., Rust, K., Blumstein, E., Krenzke, T., and Legum, S. (2001). *The 1998 High School Transcript Study User's Guide and Technical Report* (NCES 2001-477). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Roey, S., Caldwell, N., Rust, K., Hicks, L., Lee, J., Perkins, R., Blumstein, E., and Brown, J. (2005). *The High School Transcript Study: The 2000 High School Transcript Study User's Guide and Technical*

Report (NCES 2005-483). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2007). *Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007-482). Washington, DC.

White, S. (1994). *Overview of NAEP Assessment Frameworks* (NCES 94-412). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Uses of Data

Phillips, G. (1993). *Interpreting NAEP Scales* (NCES 93-421). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Vanneman, A. (1997). *Focus on NAEP: New Software Makes NAEP Data User Friendly* (NCES 97-045). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Vanneman, A., Hamilton, L., Baldwin Anderson, J., and Rahman, T. (2009). *Achievement Gaps: How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress*, (NCES 2009-455). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Survey Design

Bay, L., Chen, L., Hanson, B.A., Happel, J., Kolen, M.J., Miller, T., Pommerich, M., Sconing, J., Wang, T., and Welch, C. (1997). *ACT's NAEP Redesign Project: Assessment Design Is the Key to Useful and Stable Assessment Results* (NCES Working Paper 97-39). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Bock, D.R., and Zimowski, M.F. (2003). *NAEP Validity Studies: Feasibility Studies of Two-Stage Testing in Large-Scale Educational Assessment: Implications for NAEP* (NCES Working Paper 2003-14). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chromy, J.R. (2003). *NAEP Validity Studies: The Effects of Finite Sampling on State Assessment Sample Requirements* (NCES Working Paper 2003-17). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Durán, R.P. (2003). *NAEP Validity Studies: Implications of Electronic Technology for the NAEP Assessment* (NCES Working Paper 2003-16). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hedges, L.V., Konstantopoulos, S., and Thoreson, A. (2003). *NAEP Validity Studies: Computer Use and Its Relation to Academic Achievement in Mathematics, Reading, and Writing* (NCES Working Paper 2003-15). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hedges, L.V., and Vevea, J.L. (2003). *AEP Validity Studies: A Study of Equating in NAEP* (NCES Working Paper 2003-13). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Jaeger, R.M. (2003). *NAEP Validity Studies: Reporting the Results of the National Assessment of Educational Progress* (NCES Working Paper 2003-11). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Jakwerth, P.R., Stancavage, R.B., and Reed, E.D. (2003). *NAEP Validity Studies: An Investigation of Why Students Do Not Respond to Questions* (NCES Working Paper 2003-12). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Johnson, E.G., Lazer, S., and O'Sullivan, C.Y. (1997). *NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress* (NCES Working Paper 97-31). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Levine, R., Huberman, M., and Buckner, K. (2002). *The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth- and Eighth-Grade Students and Teachers to Questionnaire Items* (NCES Working Paper 2002-06). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

McLaughlin, D. (1997). *Can State Assessment Data Be Used to Reduce State NAEP Sample Sizes?* (NCES Working Paper 97-29). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Mullis, V. (2003). *NAEP Validity Studies: Optimizing State NAEP: Issues and Possible Improvements* (NCES Working Paper 2003-09). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Niemi, R.G. (1997). *Innovative Solutions to Intractable Large-Scale Assessment (Problem 2: Background Questionnaires)* (NCES Working Paper 97-32). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Olson, J.F., and Goldstein, A.A. (1997). *The Inclusion of Students With Disabilities and Limited English Proficient Students in Large-Scale Assessments: A Summary of Recent Progress* (NCES 97-482). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Owen, E.H. (1992). *Exploring New Methods for Collecting Students' School-Based Writing* (NCES 92-065). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Patz, R.J., Wilson, M., and Hoskens, M. (1997). *Optimal Rating Procedures and Methodology for NAEP Open-Ended Items* (NCES Working Paper 97-37). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Pearson, D.P., and Garavaglia, D.R. (2003). *NAEP Validity Studies: Improving the Information Value of Performance Items in Large-Scale Assessments* (NCES Working Paper 2003-08). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

Stancavage, F.B. (2003) *NAEP Validity Studies: An Agenda for NAEP Validity Research Validity Research* (NCES Working Paper 2003-07). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Weston, T.J. (2003). *NAEP Validity Studies: The Validity of Oral Accommodation in Testing* (NCES Working Paper 2003-06). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Bandeira de Mello, V., Blankenship, C., and McLaughlin, D.H. (2009). *Mapping State Proficiency Standards Onto NAEP Scales: 2005-2007* (NCES 2010-456). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Binkley, M., and Kelly, D. (2003). *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments* (NCES Working Paper 2003-10). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

DeVito, P.J., and Koenig, J.A. (2001). *NAEP Reporting Practices: Investigating District-Level and Market-Basket Reporting.* Washington, DC: National Research Council.

Hoffman, G.R., Becker, D.E., and Wise, L. (2003). *NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware* (NCES Working Paper 2003-19). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ingels, S., and Taylor, J. (1995). *National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data* (NCES Working Paper 95-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Johnson, E.G. (1998). *Linking the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): A Technical Report* (NCES 98-499). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Johnson, E., Cohen, J., Chen, W.H., Jiang, T., and Zhang, Y. (2005). *2000 NAEP–1999 TIMSS Linking Report* (NCES Working Paper 2005-01). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kitmitto, S., and Bandeira de Mello, V., (2008). *Measuring the Status and Change of NAEP State Inclusion Rates for Students with Disabilities* (NCES 2009–453). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Neidorf, T.S., Binkley, M., Gattis, K., and Nohara, D. (2006). *Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments* (NCES 2006-029). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Neidorf, T.S., Binkley, M., and Stephens, M. (2006). *Comparing Science Content in the National Assessment of Educational Progress (NAEP) 2000 and Trends in International Mathematics and Science Study (TIMSS) 2003 Assessments* (NCES 2006-026). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Pellegrino, J.W., Jones, L.R., and Mitchell, K.J. (1999). *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. Washington, DC: National Research Council.

Phillips, G. W (2009), *The Second Derivative: International Benchmarks in Mathematics for U.S. States and School Districts*. Washington, DC: American Institutes for Research.

Raju, N.S., Pellegrino, J.W., Bertenthal, M.W., Mitchell, K.J., and Jones, L.R. (2000). *Grading the Nation's Report Card: Research From the Evaluation of NAEP*. Washington, DC: National Research Council.

Sedlacek, D.A. (1995). *Model-Based Methods for Analysis of Data From 1990 NAEP Trial State Assessment* (NCES 95-696). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Sedlacek, D.A. (1995). *Use of Person-Fit Statistics in Reporting and Analyzing National Assessment of Educational Progress Results* (NCES 95-713). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Sedlacek, D.A. (1995). *Using HLM and NAEP Data to Explore School Correlates of 1990 Mathematics and Geometry Achievement in Grades 4, 8, 12—Methodology and Results* (NCES 95-697). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Chapter 19: National Adult Literacy Survey (NALS)

1. OVERVIEW

PERIODIC SURVEY OF A SAMPLE OF ADULTS LIVING IN HOUSEHOLDS OR PRISONS

Assesses literacy skills:

- Prose
- Document
- Quantitative

Collects background data:

- Demographics
- Education
- Labor market experiences
- Income
- Activities

The National Adult Literacy Survey (NALS) was initiated to fill the need for accurate and detailed information on the English literacy skills of America's adults. In accordance with a congressional mandate, it provided the most detailed portrait that has ever been available in the 1990s on the condition of literacy in this nation.

The 1992 NALS is the third assessment of adult literacy funded by the federal government and conducted by the Educational Testing Service (ETS). The two previous efforts were (1) the 1985 Young Adult Literacy Assessment, funded as an adjunct to the National Assessment of Educational Progress (NAEP)—see chapter 18); and (2) the Department of Labor's 1990 Workplace Literacy Survey. Building on these two earlier surveys, literacy for NALS is defined along three dimensions—prose, document, and quantitative—designed to capture an ordered set of information-processing skills and strategies that adults use to accomplish a diverse range of literacy tasks encountered in everyday life. The background data collected in NALS provide a context for understanding the ways in which various characteristics are associated with demonstrated literacy skills.

NALS is the first national study of literacy for *all* adults since the Adult Performance Level Surveys conducted in the early 1970s. It is also the first in-person literacy assessment involving the prison population. A second adult literacy survey, the National Assessment of Adult Literacy (NAAL), was conducted in 2003.

Purpose

To (1) evaluate the English language literacy skills of adults (16 years and older) living in households or prisons in the United States; (2) relate the literacy skills of the nation's adults to a variety of demographic characteristics and explanatory variables; and (3) compare the results with those from the 1985 Young Adult Literacy Assessment and the 1990 Workplace Literacy Survey.

Components

The 1992 survey consisted of one component that was administered to three different representative samples: a national household sample; supplemental state household samples for 12 states (California, Florida, Illinois, Indiana, Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, and Washington); and a national sample of federal and state prison inmates. Responses from the national, state, and prison samples were combined to yield the best possible performance estimates.

National Adult Literacy Survey. The 1992 survey assessed the literacy skills of a representative sample of the U.S. adult population using simulations of three kinds of literacy tasks that adults would ordinarily encounter in daily life(prose, document, and quantitative literacy). The data were collected through in-person

interviews with adults who were living in households or in federal or state prisons. Adults were defined as individuals 16 years or older for the national and prison samples, and 16 to 64 years of age for the state samples. In addition to the cognitive tasks, the personal interview gathered information on demographic characteristics, language background, educational background, reading practices, and labor market experiences. To ensure comparability across all samples, the literacy tasks assessed were the same for all three samples. Background data varied somewhat between the household and prison samples—labor force questions were irrelevant to prisoners, and questions about criminal behavior and sentences were relevant only to prisoners.

Literacy Assessment. The pool of literacy tasks used to measure adult proficiencies consisted of 165 literacy questions—41 prose, 81 document, and 43 quantitative. To ensure that valid comparisons could be made by linking the scales to those of the 1985 Young Adult Literacy Assessment, 85 tasks from that survey were included in the 1992 survey. An additional 80 new tasks were developed specifically to complement and enhance the original 85 tasks. The literacy tasks administered in NALS varied widely in terms of materials and content. The six major context/content areas were home and family; health and safety; community and citizenship; consumer electronics; work; and leisure and recreation. Each adult was given a subset (about 45) of the total pool of assessment tasks to complete. Each of the tasks extended over a range of difficulty on the three literacy scales. The new tasks were designed to simulate the way in which people use various types of materials and to require different strategies for successful performance.

The responses to the literacy assessment were pooled and reported by proficiency scores, ranging from 0 to 500, on three separate scales, one each for prose, document, and quantitative literacy. By examining the overall characteristics of individuals who performed at each literacy level on each scale, it is possible to identify factors associated with higher or lower proficiency in reading and using prose, document, and quantitative materials.

Background Information. Background information collected for the state and household samples included data on *background and demographics*—country of birth, languages spoken or read, access to reading materials, size of household, educational attainment of parents, age, race/ethnicity, and marital status; *education*—highest grade completed in school, current aspirations, participation in adult education classes, and education received outside the country; *labor market experiences*—employment status, recent labor market experiences, and occupation; *income*—personal and household; and *activities*—voting behavior, hours spent watching television, frequency and content of newspaper reading, and use of literacy skills for work and leisure. Respondents from each of the 12 participating states were also asked state-specific questions.

To address issues of particular relevance to the prison population, a separate background questionnaire was developed for the prison sample. This instrument drew questions from the 1991 Survey of Inmates of State Correctional Facilities, sponsored by the Department of Justice's Bureau of Justice Statistics. The background questionnaire for the prison population addressed the following major topics: general and language background; educational background and experience; current offenses and criminal history; prison work assignments and labor force participation prior to incarceration; literacy activities and collaboration; and demographic information.

Periodicity

NALS was conducted in 1992. NAAL, a continuation of NALS, was conducted in 2003.

2. USES OF DATA

Results from NALS provide a detailed portrait on the condition of literacy in this nation. NALS data provide vital information to policymakers, business and labor leaders, researchers, and citizens. The survey results can be used to

- describe the levels of literacy demonstrated by the adult population as a whole and by adults in various subgroups (e.g., those targeted as at risk, prison inmates, and older adults);
- characterize adults' literacy skills in terms of demographic and background information (e.g., reading characteristics, education, and employment experiences);
- profile the literacy skills of the nation's workforce;
- compare assessment results from the current study with those from the 1985 Young Adult Literacy Assessment;
- interpret the findings in light of information-processing skills and strategies, so as to inform

curriculum decisions concerning adult education and training; and

- increase understanding of the skills and knowledge associated with living in a technological society.

3. KEY CONCEPTS

Some of the key concepts related to the literacy assessment are described below. See the NALS Electronic Codebook or appendices of NALS reports for lists and descriptions of variables.

Literacy. The ability to use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential. This definition goes beyond simply decoding and comprehending text to include a broad range of information-processing skills that adults use in accomplishing the range of tasks associated with work, home, and community contexts.

Prose Literacy. The ability to locate information contained in expository or narrative prose in the presence of related but unnecessary information, find all of the relevant information, integrate information from various parts of a passage of text, and write new information related to the text. Expository prose consists of printed information in the form of connected sentences and longer passages that define, describe, or inform, such as newspaper stories or written instructions. Narrative prose tells a story, but is less frequently used by adults in everyday life than by school children, and did not occur as often in the text presented in NALS as prose literacy tasks. Prose varies in its length, density, and structure.

Document Literacy. The ability to locate information in documents, repeat the search as many times as needed to find all the information, integrate information from various parts of a document, and write new information as requested in appropriate places in a document, while screening out related but inappropriate information. Documents differ from prose text in that they are more highly structured. Documents consist of structured prose and quantitative information in complex arrays arranged in rows and columns, such as tables, data forms, and lists (simple, nested, intersected, or combined); in hierarchical structures, such as tables of contents or indexes; or in two-dimensional visual displays of quantitative information, such as graphs, charts, and maps.

Quantitative Literacy. The ability to use quantitative information contained in prose or documents (specifically the ability to locate quantities while screening out related but unneeded information), repeat the search as many times as needed to find all the numbers, integrate information from various parts of a text or document, infer the necessary arithmetic operation(s), and perform arithmetic operation(s). Quantities can be located in either prose texts or in documents. Quantitative information may be displayed visually in graphs, maps, or charts, or it may be displayed numerically using whole numbers, fractions, decimals, percentages, or time units (hours and minutes).

Literacy Scales. Three scales used to report the results for prose, document, and quantitative literacy. These scales, each ranging from 0 to 500, are based on those established for the 1985 Young Adult Literacy Assessment. The scores on each scale represent degrees of proficiency along that particular dimension of literacy. The literacy tasks administered in the 1992 survey varied widely in terms of materials, content, and task requirements, and thus in difficulty. A careful analysis of the range of tasks along each scale provides clear evidence of an ordered set of information-processing skills and strategies along each scale. To capture this ordering, each scale was divided into five levels that reflect this progression of information-processing skills and strategies: Level 1 (0 to 225), Level 2 (226 to 275), Level 3 (276 to 325), Level 4 (326 to 375), and Level 5 (376 to 500). Level 1 comprised those adults who could consistently succeed with Level 1 literacy tasks but not with Level 2 tasks, as well as those who could not consistently succeed with Level 1 tasks and those who were not literate enough in English to take the test at all. Adults in Levels 2 through 4 were consistently able to succeed with tasks at their level but not with the next more difficult level of tasks. Adults in Level 5 were consistently able to succeed with Level 5 tasks.

Succeed Consistently. Indicates that a person at or above a given level of literacy has at least an 80 percent chance of correctly responding to a particular task. This 80 percent criterion is more stringent than the 65 percent standard used in NAEP (see chapter 18) for measuring what school children know and can do.

4. SURVEY DESIGN

The 1992 NALS was designed and administered by ETS. A subcontract was awarded to Westat, Inc., for sampling and field data collection. A committee of

experts from business and industry, labor, government, research, and adult education worked with the ETS staff to develop the definition of literacy that underlies NALS, as well as to prepare the assessment objectives that guided the selection and construction of assessment tasks. In addition to this Literacy Definition Committee, a Technical Review Committee was formed to help ensure the soundness of the assessment design, the quality of the data collected, the integrity of the analyses conducted, and the appropriateness of the interpretations of the final results. The prison survey was developed in consultation with the Bureau of Justice Statistics and the Federal Bureau of Prisons. The survey design for the 1992 survey is described below.

Target Population

The target population for the national household sample consisted of adults 16 years and older in the 50 states and the District of Columbia who, at the time of the survey, resided in private households or college dormitories. The target population for the supplemental state household sample consisted of individuals 16 to 64 years of age who, at the time of the survey, resided in private households or college dormitories in the participating state (California, Florida, Illinois, Indiana, Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, or Washington). Individuals residing in other institutions—nursing homes, group homes, or psychiatric facilities—were not included in the household samples. The target population for the prison sample consisted of adults 16 years or older who were in state or federal prisons at the time of the survey; those held in local jails, community-based facilities, or other types of institutions were not included.

Sample Design

Because this 1992 survey was designed to provide data representative at the national level (including prison inmates) and at the state level for participating states, it included three different samples: a national household sample, supplemental state household samples for 12 states, and a supplemental national sample of state and federal prison inmates.

Household Samples. The sample design for the national and state household samples involved a four-stage stratified area sample: (1) the selection of primary sampling units (PSUs) consisting of counties or contiguous groups of counties; (2) the selection of segments (within the selected PSUs) consisting of census blocks or groups of contiguous census blocks; (3) the selection of households within the segmented samples; and (4) the selection of age-eligible individuals within each selected household. The sample design requirements called for an average cluster size of seven interviews (i.e., seven completed background questionnaires per segment). In addition, a reserve sample at the household level of approximately 5 percent of the size of the main sample was selected and set aside in case of shortfalls due to unexpectedly high vacancy and nonresponse rates.

One national area sample was drawn for the national household sample, and 12 independent state-specific area samples were drawn from the 12 states participating in the supplemental state samples. The sample designs used for all 13 samples were similar, with one major difference. In the national sample, Black and Hispanic respondents were sampled at about double the rate of the remainder of the population to assure reliable estimates of their literacy proficiencies, whereas the state samples used no oversampling.

The first stage of sampling involved the selection of PSUs. A national sampling frame of 1,400 PSUs was constructed primarily from 1990 census data stratified on the basis of region, metropolitan status, percent Black, percent Hispanic, and whenever possible, per capita income. Using this frame, 101 PSUs were selected for the national sample. The national frame of PSUs (subdivided at state boundaries, if needed) was used to construct individual state frames for the supplemental state sample; a sample of 8 to 12 PSUs was selected within each of the given states. All PSUs were selected with probability proportional to the PSU's 1990 population.

The second stage of sampling involved the selection of segments within the selected PSUs. The Bureau of the Census's Topologically Integrated Geographical Encoding and Referencing (TIGER) System File was used for the production of segment maps. The segments were selected with probability proportional to size, where the measure of size for a segment was a function of the number of year-round housing units within the segment. The oversampling of Black and Hispanic respondents for the national sample was carried out at the segment level, where segments were classified either as having a high percentage of the Black or Hispanic population (more than 25 percent) or as not having a high percentage.

The third stage of sampling involved the selection of households within the segmented samples. Westat field staff visited all selected segments in the fall of 1991 and prepared lists of all housing units within the boundaries of each segment as determined by the 1990 census block maps. The lists were used to construct the sampling frame for households. Households were selected with equal probability within each segment,

except for White, non-Hispanic households in segments with a high percentage of the Black or Hispanic population (over 25 percent) in the national sample, which were subsampled so that the sampling rates for White, non-Hispanic respondents would be about the same overall.

The fourth stage of sampling involved the selection of one or two adults within each selected household during the data collection phase of the survey. One person was selected at random from households with fewer than four eligible members; two persons were selected from households with four or more eligible members. Using a screener, the interviewer constructed a list of age-eligible household members (16 and older for the national sample, 16 to 64 for the state sample) for each selected household. The interviewers, who were instructed to list the eligible household members in descending order by age, then identified one or two household members to interview, based on computer-generated sampling messages that were attached to each questionnaire in advance.

Prison Sample. There were two stages of selection for the prison sample. The first stage involved the selection of state or federal correctional facilities. The sampling frame for the correctional facilities was based on the 1990 census of federal and state prisons, updated in mid-1991. The facility frame was stratified prior to sample selection on the basis of type of facility (federal or state prison), region of country, inmate gender composition, and type of security. A sample of 88 facilities and a reserve sample of 8 facilities was then drawn from the frame based on probability proportional to size, where the measure of size for a given facility was equal to the inmate population. The second stage of sampling involved the selection of inmates within each selected facility, using a list of names obtained from the facility administrators. An average of 12 inmates were selected from each facility based on a probability inversely proportional to their facility's inmate population (up to a maximum of 22 interviews in a facility), so that the product of the first- and second-stage probabilities would be constant.

Assessment Design

Building on the 1985 Young Adult Literacy Assessment and the 1991 Workplace Literacy Survey, the NALS Technical Committee adopted the definition of literacy and the literacy scales—prose, document, and quantitative—used in the previous surveys. The materials were selected to represent a variety of contexts and contents: home and family; health and safety; community and citizenship; consumer electronics; work; and leisure and recreation.

BIB Spiraling. The survey design gave each respondent a subset of the total pool of literacy tasks, while at the same time ensuring that each of the 165 tasks was administered to a nationally representative sample of the adult population. The design most suitable for this purpose is a variant of standard matrix sampling called balanced incomplete block (BIB) design.

Literacy tasks were assigned to blocks or sections that could be completed in about 15 minutes, and these blocks were then compiled into booklets so that each block appeared in each position (first, middle, and last) and each block was paired with every other block. Thirteen blocks of simulation tasks were assembled into 26 unique booklets, each of which contained four blocks of tasks: the core (the same for all exercise booklets) and three cognitive blocks. Each booklet could be completed in about 45 minutes.

Pretests. A field test of the national household sample was conducted in the spring of 1991 using a sample of 2,000 adults drawn from 16 PSUs. The purposes of the field test were to evaluate the impact of incentives on response rates, performance, and survey costs; to evaluate newly developed literacy exercises for item bias and testing time; and to evaluate the administration and appropriateness of the background questions. As a result of the field test, some of the literacy tasks and their scoring guides were revised or dropped from the final assessment.

For the prison sample, a small pretest was conducted at the Roxbury Correctional Institution in Hagerstown, Maryland. This pretest was designed to evaluate the ease of administration of the survey instruments, survey administration time, within-facility procedures, and inmate reaction to the survey. The pretest demonstrated that several changes to the background questionnaire would facilitate administration. Administrative procedures were also refined to reflect lessons learned during the pretest.

Data Collection and Processing

The survey data were collected through in-person household or prison interviews during the first 8 months of 1992. As field operations were completed, the data were shipped to ETS for processing. Further description follows.

Reference Dates. Respondents answered the employment status and weekly wages questions for the week before the survey was administered.

Data Collection. During January and February of 1992, field interviewers, supervisors, and editors received

extensive training both in general and survey-specific interview techniques. The NALS field period began in February 1992, immediately following the completion of the first interviewer training sessions, and lasted 28 weeks, until the end of August. All three survey sample groups were worked simultaneously (except for the state of Florida, where data were not collected until 1993). Except for a small, experimental "no incentive" group, all household participants who completed as much of the assessment as their skills allowed received $20 for their time. More than 400 trained interviewers visited about 44,000 households to select and interview almost 31,000 adults. In addition, over 1,147 prison inmates at 87 facilities were interviewed.

Each survey participant was asked to spend approximately one hour responding to survey questions and tasks. Data collection instruments included the screener (designed to enumerate household members and select survey respondents), the background questionnaire, and the literacy exercise booklets. Answering the screener and background questionnaire required no reading or writing skills; to ensure standardized administration, the questions on each were read to respondents in English or Spanish and the answers recorded by the assessment interviewer. Each of the exercise booklets had a corresponding interview guide, with specific instructions to the interviewer for directing the exercise booklet. Reading and writing skills in the English language were required to complete the exercise booklet. When a sampled respondent did not complete any or all of the survey instruments, the interviewer was required to complete a noninterview report form. Field supervisors reviewed the noninterview forms to determine the case's potential for conversion, and the data collected on the form were processed for nonresponse analysis.

Following the completion of an interview, interviewers edited all materials for legibility and completeness. The interviewers sent their completed work to their regional supervisors for a complete edit of the instruments, quality control procedures, and any required data retrieval. As these tasks were completed, the cases were shipped to ETS for processing.

During the data collection process, two special quality control procedures were implemented to identify any households or dwellings missed during the listing phase: the missing structure procedure and the missed dwelling unit procedure. These procedures were used to give these missed structures and dwelling units a chance of selection at time of data collection.

The field effort occurred in three overlapping stages:

(1) *Initial Phase.* Each area segment was assigned by the regional supervisor to an interviewer, who followed certain rules in making a prescribed number of calls (a maximum of four was used) to every sampled dwelling in the segment.

(2) *Reassignment Phase.* Cases that did not result in completed interviews during the initial phase were reviewed by the regional supervisor, and a subset was selected for reassignment to another interviewer in the same PSU or an interviewer from a nearby PSU.

(3) *Special Nonresponse Conversion Phase.* The home office assembled a special traveling team of the most experienced or productive interviewers to perform a nonresponse conversion effort, under the supervision of a subset of the field supervisors.

Data Processing. Coding and scoring staff underwent intensive training prior to the actual coding and scoring. A scoring supervisor monitored both the coding of the questionnaires and the scoring of the exercise booklets. The background questionnaire was designed to be read by a computerized scanning device. Nearly all the simulation tasks contained in the exercise booklet were open-ended; with scoring guides as examples, responses to these items were classified as correct, incorrect, or omitted by trained readers. Responses from the screener and scores from the exercise booklets were transferred to scannable answer sheets. Each survey instrument's scannable forms were batched and sent to the scanning department at regular intervals. As the different instruments were processed, the data were transferred to a database on the main ETS computer for editing.

Editing. Several quality control procedures related to data collection were used during the field operation: an interviewer field edit, a complete edit of all documents by a trained field editor, validation of 10 percent of each interviewer's closeout work, and field observation of both supervisors and interviewers. Additional edits were done during data processing. These included an assessment of the internal logic and consistency of the data received. Discrepancies were corrected whenever possible. The background questionnaires were also checked to make sure that the skip patterns had been followed and all data errors were resolved. In addition, a random set of exercise booklets was selected to provide an additional check on the accuracy of transferring information from booklets and answer sheets to the database.

Estimation Methods

Weighting was used in the 1992 NALS, prior to the calculation of base weights. Responses to the literacy tasks were scored using item response theory (IRT) scaling. A multiple imputation procedure based on plausible values methodology was used to estimate the literacy proficiencies of individuals who completed literacy tasks. An innovative approach was implemented to impute missing cognitive data in order to minimize distortions in the population proficiency estimates due to nonresponse to the literacy booklet.

Weighting. Full sample and replicate weights were calculated for survey respondents who completed the exercise booklet; those who could not start the exercises because of a language barrier, a physical or mental barrier, or a reading or writing barrier; and those who refused to complete the exercises but had completed background questionnaires. Demographic variables critical to the weighting were recoded and imputed, if necessary, prior to the calculation of base weights (see “Imputation” below). Separate sets of weights were computed for the incentive and “no incentive” samples.

Household samples. A base weight was computed for each eligible record. The base weight initially was computed as the reciprocal of the product of probabilities of selection for a respondent at the PSU, segment, dwelling unit, and person levels. The final base weight included adjustments to reflect the selection of the reserve sample, the selection of missed dwelling units, and the chunking process conducted during the listing of the segments; and to account for the subsample of segments assigned to the “no incentive” experiment and the subsampling of respondents within households. The base weights for each sample were then poststratified to known 1990 census population totals, adjusted for undercount. This first-level stratification provided sampling weights with lower variation and adjusted for nonresponse. State records were poststratified separately from national records to provide a common base for applying composite weighting factors; population totals were calculated separately for each distinct group.

Composite weights were developed so that NALS data could be used to produce both state and national statistics. For the household samples, a composite weight was computed as the product of the poststratified base weight and a compositing factor that combined the national and state sample data in an optimal manner, considering the differences in sample design, sample size, and sampling error between the two sampled groups. Up to four different compositing factors were used in each of the 11 participating states, and a pseudo-factor (equal to 1) was used for all persons 65 and older and for all national sample records from outside the 11 participating states.

To compute the final sample weights, the composite weights were adjusted to known 1990 census counts (adjusted for undercount), using a process called the poststratification raking ratio adjustment. The cells used for raking were defined to the finest combination of age, race/ethnicity, sex, education, and geographic indicators (e.g., Metropolitan Statistical Area [MSA] vs. non-MSA) that the data would allow. Raking adjustment factors were calculated separately for each of the state samples and then for the remainder of the United States.

The above steps used to create the final sample weights were repeated for 60 strategically constructed subsets of the household sample to create a set of replicate weights to be used for variance estimation using the jackknife method.

Prison sample. Base weights for the prison respondents were constructed to be equal to the reciprocal of the product of the selection probabilities for the facility and the inmate within the facility. These weights were then nonresponse-adjusted to reflect both facility and inmate nonresponse. To compute the final sample weights, the resulting nonresponse-adjusted weights were then raked to agree with independent estimates for certain subgroups of the prison population. The above procedures were repeated for 45 strategically constructed subsets of the prison sample to create a set of replicate weights to be used for variance estimation using the jackknife method.

Scaling. Since NALS used a variant of matrix sampling and since different respondents received different sets of tasks, it would be inappropriate to report its results using conventional scoring methods based on the number of correct responses. The literacy assessment results are reported using IRT scaling, which assumes some uniformity in response patterns when items require similar skills. Such uniformity can be used to characterize both examinees and items in terms of a common scale attached to the skills, even when all examinees do not take identical sets of items. Comparisons of items and examinees can then be made in reference to a scale, rather than to the percent correct. IRT scaling also allows the distributions of examinee groups to be compared.

The results of the 1992 literacy assessment are reported on three scales (prose, document, and quantitative) that were established for the 1985 Young Adult Literacy Assessment. Separate IRT linking and scaling were

carried out for each of the three domains, using the three-parameter logistic (3PL) scaling model from item response theory. This is a mathematical model for estimating the probability that a particular person will respond correctly to a particular item from a single domain of items. The probability is given as a function of a parameter characterizing the proficiency of that person and three parameters characterizing the properties of that item. Item parameters needed for the 3PL scaling model were estimated by linking each of the literacy scales used in the 1992 survey to the 1985 Young Adult Literacy Assessment scales.

Imputation. Imputation was performed prior to weighting on missing demographic items considered critical to weighting. Literacy proficiencies of respondents were estimated using a multiple imputation procedure based on plausible values methodology. Missing cognitive data were also imputed.

Demographic data. Demographic variables critical to the weighting (race/ethnicity of the head of household; sex, age, race/ethnicity, and education of the respondent) were recoded and collapsed to required levels, and imputed, if necessary, prior to the calculation of base weights. Data from the background questionnaire were preferred for all items except race/ethnicity of the head of household, which was collected in the screener. For the few cases in which the background questionnaire measure was missing, the screener measure was generally available and was used as a direct substitute. The amount of missing data remaining after substitution was small, making the imputation task fairly straightforward. A standard (random within class) hot-deck imputation procedure was performed for particular combinations of fields that were missing. Imputation flags were created for each of the five critical fields to indicate whether data were originally reported or were based on substitution or imputation. The imputed values were used only for the sample weighting process.

Literacy proficiency estimation (plausible values). A multiple imputation procedure based on plausible values methodology was used to estimate respondents' literacy proficiency in the 1992 NALS. When analyzing the distribution of proficiencies in a group of persons, more efficient estimates can be obtained from a sample design similar to that used in this 1992 survey. Such designs solicit relatively few cognitive responses from each sampled respondent, but maintain a wide range of content representation when responses are summed for all respondents.

In the 1992 survey, all proficiency data were based on two types of information: responses to the background questions and responses to the cognitive items. As an intermediate step, a functional relationship between the two sets of information was calculated for the total sample, and this function was used to obtain unbiased proficiency estimates for population groups with reduced error variance. Possible values for a respondent's proficiency were sampled from a posterior distribution that is the product of two functions: the conditional distribution of proficiency given the pattern of background variables and the likelihood function of proficiency given the pattern of responses to the cognitive items. Since exact matches of background responses are quite rare, NALS used more than 200 principal components to summarize the background information, capturing more than 99 percent of the variance. More detailed information on the plausible values methodology used in the 1992 survey is available in the *Technical Report and Data File User's Manual for the 1992 National Adult Literacy Survey* (Kirsch et al. 2000).

Cognitive data. New procedures were implemented in the 1992 NALS to minimize distortions in the population proficiency estimates due to nonresponse to the literacy booklets. When a sampled individual decided to stop the assessment (answered less than five literacy items per scale), the interviewer used a standardized nonresponse coding procedure to record the reason why the person was stopping. This information was used to classify nonrespondents into two groups: (1) those who stopped the assessment for literacy-related reasons (e.g., language difficulty, mental disability, or reading difficulty not related to a physical disability); and (2) those who stopped for reasons unrelated to literacy (e.g., physical disability or refusal). About half of the individuals did not complete the assessment for reasons related to their literacy skills; the other respondents gave no reason for stopping or gave reasons unrelated to their literacy.

To represent the range of implied causes of missing literacy responses, the imputation procedure selected relied on background variables and self-reported reasons for nonresponse, in addition to the functional relationship between background variables and proficiency scores for the total population. It treated "consecutively missing" data from the literacy booklet instrument differently depending on whether the nonrespondents' reasons were related or unrelated to their literacy skills: (1) those who gave literacy-related reasons were treated as wrong answers, based on the assumption that they could not have correctly completed the literacy tasks, whereas (2) those who gave no reason or cited reasons unrelated to literacy

skills for not completing the assessment were essentially ignored (considered not reached), since it could not be assumed that their answers would have been either correct or incorrect. The proficiencies of such respondents were inferred from the proficiencies of other adults with similar characteristics using the plausible values methodology described above.

Future Plans

A second survey, NAAL, was conducted in 2003. Currently, there are no plans to administer another measure of adult literacy.

5. DATA QUALITY AND COMPARABILITY

The NALS sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. In addition, NCES conducted special evaluation studies to examine issues related to the quality of NALS. These studies included (1) a study of the role of incentives in literacy survey research; (2) an evaluation of its sample design and composite estimation; and (3) an evaluation of the construct validity of the adult literacy scales.

Sampling Error

In the 1992 survey, the use of a complex sample design, adjustments for nonresponse, and poststratification procedures resulted in dependence among the observations. Therefore, a jackknife replication method was used to estimate the sampling variance. The mean square error of replicate estimates around their corresponding full sample estimate provides an estimate of the sampling variance of the statistic of interest. The replication scheme was designed to produce stable estimates of standard errors for national and prison estimates as well as for the 12 individual states.

The advantage of compositing the national and state samples during sample weighting was the increased sample size, which improved the precision of both the state and national estimates. However, biases could be present because the national PSU sample strata were not designed to maximize the efficiency of state-level estimates.

Nonsampling Error

The major source of nonsampling error in the 1992 NALS was nonresponse error; special procedures were developed to minimize potential nonresponse bias based on how much of the survey the respondent completed. Other possible sources of nonsampling error were random measurement error and systematic error due to interviewers, coders, or scorers.

Coverage Error. Coverage error could result from either the sampling frame of households or prisons being incomplete or from a household's or prison's failure to include all adults 16 years and older on the lists from which the sampled respondents were drawn. Special procedures and edits were built into NALS to review both listers' and interviewers' ongoing work and to give any missed structures and/or dwelling units a chance of selection at data collection. However, just as all other household personal interview surveys have persistent undercoverage problems, the 1992 survey had problems in population coverage due to interviewers not gaining access to households in dangerous neighborhoods, locked residential apartment buildings, and gated communities.

Nonresponse Error.
Unit nonresponse. Since three survey instruments—screener, background questionnaire, and exercise booklet—were required for the administration of the survey, it was possible for a household or respondent to refuse to participate at the time of the administration of any one of these instruments. Because the screener and background questionnaire were read to the survey participants in English or Spanish, but the exercise booklet required reading and writing in the English language, it was possible to complete the screener or background questionnaire but not the exercise booklet, and vice versa. Thus, response rates were calculated for each of the three instruments for the household samples (see table 12). For the prison sample, there were only two points at which a respondent could not respond—at the administration of the background questionnaire or the exercise booklet.

The response rate to the background questionnaire was 80.5 percent. For the household samples, the response rates exclude individuals who were not paid incentives. Also excluded are the respondents to the Florida state survey, which had a delayed administration.

The combined national and state household target sample in the 1992 NALS included 43,780 representative housing units, of which 5,410 were vacant. Approximately 89 percent of the occupied households completed a screener.

The household sample screening effort identified a total of 30,810 eligible respondents, of whom 24,940 (81.0 percent unweighted) completed the background questionnaire. For the prison sample, 87 of the 88 sampled facilities participated in the survey. Of the 1,340 inmates selected, 1,150 (85.6 percent

unweighted) completed the background questionnaire. For the occupied households, “refusal or breakoff” was the most common explanation for nonresponse to the screener and background questionnaire. The second most common explanation was “not at home after maximum number of calls.” Nonresponse also resulted from language, physical, and mental problems. Housing units or individuals who refused to participate before any information was collected about them, or who did not answer a sufficient number of background questions, were never incorporated into the database. Because these individuals were unlikely to know that the survey intended to assess their literacy, it was assumed that their reason for not completing the survey was not related to their level of literacy.

Literacy assessment booklets were considered complete if at least five items were answered on each scale. A total of 24,940 household sample members were classified as eligible for the exercise booklet. Of these, 88.6 percent completed the booklet and another 6.1 percent partially completed it. Of the 1,150 eligibles in the prison sample, 86.8 percent completed the booklet and another 9.3 percent partially completed it.

There were reasons to believe that the literacy performance data were missing more often for adults with lower levels of literacy than for adults with higher levels. Field-test evidence and experience with surveys indicated that adults with lower levels of literacy were more likely than adults with higher proficiencies either to decline to respond to the survey at all or to begin the assessment but not complete it. Ignoring this pattern of missing data would have resulted in overestimating the literacy skills of adults in the United States. Therefore, to minimize bias in the proficiency estimates due to nonresponse to the literacy assessment, special procedures were developed to impute the literacy proficiencies of nonrespondents who completed fewer than five literacy tasks.

Item nonresponse. For each background questionnaire, staff verified that certain questions providing critical information for weighting and data analyses had been answered, namely, education level, employment status, parents’ level of education, race, and sex. If a response was missing, the case was returned to the field for data retrieval. Therefore, item response rates for completed background questionnaires were quite high, although they varied by type of question. Questions asking country of origin (first question in the booklet) and sex (last question in the booklet) had nearly 100 percent response rates, indicating that most respondents attempted to complete the entire questionnaire. Response rates were lower, however, for questions about income and educational background.

Table 12. Weighted and unweighted response rates for all sample types in the National Adult Literacy Survey, by survey component: 1992

Component	Weighted (percent)	Unweighted (percent)
Screener	—	89.1
Background questionnaire	80.5	81.0
Exercise booklet	95.9	95.9

— Not available.
NOTE: The weighted response rates were calculated by applying the sampling weight to each individual to account for his or her probability of selection into the sample. Weighted response rates were computed only for screened households (the probability of selection is not known for persons in households that were not screened).
SOURCE: Kirsch, I.S., Yamamoto, K., Norris, N., Rock, D., Jungeblut, A., O’Reilly, P., Campbell, A., Jenkins, L., Kolstad, A., Berlin, M., Mohadjer, L., Waksberg, J., Goksel, H., Burke, J., Rieger, S., Green, J., Klein, M., Mosenthal, P., and Baldi, S. (2000). *Technical Report and Data File User’s Manual for the 1992 National Adult Literacy Survey* (NCES 2001-457). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

The electronic codebook provides counts of item nonresponse. These, however, have to be considered in terms of the number of adults that were offered each task, because a great deal of the missing data is missing by design.

Measurement Error. All background questions and literacy tasks underwent extensive review by subject area and measurement specialists, as well as scrutiny to eliminate any bias or lack of sensitivity to particular groups. Special care was taken to include materials and tasks that were relevant to adults of widely varying ages. During the test development stage, the tasks were submitted to test specialists for review, part of which involved checking the accuracy and completeness of the scoring guide. After preliminary versions of the assessment instruments were developed and after the field test was conducted, the literacy tasks were closely analyzed for bias or “differential item functioning.” The goal was to identify any assessment tasks that were likely to underestimate the proficiencies of a particular subpopulation, whether it be older adults, females, or Black or Hispanic adults. Any assessment item that appeared to be biased against a subgroup was excluded from the final survey. The coding and scoring guides also underwent further revisions after the first responses were received from the main data collection.

Interviewer error checks. Several quality control procedures related to data collection were used during the field operation: an interviewer field edit, a complete edit of all documents by a trained field editor, validation of 10 percent of each interviewer's closeout work, and field observation of both supervisors and interviewers.

Coding/scoring error checks. In order to monitor the accuracy of coding, the questions dealing with country of birth, language, wages, and date of birth were checked in 10 percent of the questionnaires by a second coder. For the industry and occupation questions, 100 percent of the questionnaires were recoded by a second coder. Twenty percent of all the exercise booklets were subjected to a reader reliability check, which entailed a scoring by a second reader. There was a high degree of reader reliability across tasks—ranging from 88.1 to 99.9 percent—with an average agreement of 97 percent. For 133 out of 165 open-ended tasks, the agreement between the two readers was above 95 percent.

Data Comparability

One of the major goals of this survey was to compare its results to the 1985 Young Adult Literacy Assessment and other large assessment studies. NALS is also comparable with NAAL, conducted in 2003, in terms of assessment scores (see chapter 20).

Comparisons with the 1985 Young Adult Literacy Assessment. Comparisons are possible because the sample design, item pool, and methodology used in the 1985 Young Adult Literacy Assessment and the 1992 survey were very similar. Literacy tasks for each survey were developed using the same definition of literacy, and a subset of identical tasks was administered in both assessments. Scoring guides were the same for both surveys. Both gave nearly identical incentive payments to participants ($15 in 1985 and $20 in 1992). The literacy scales used in the two surveys were linked so that the scores could be reported on a common scale.

Nevertheless, there were some differences in procedures for the two surveys. For example, missing responses to the literacy tasks were handled differently. In the 1985 Young Adult Literacy Assessment, individuals who could not answer six core literacy tasks and those who spoke only Spanish were excluded from the analyses. In the 1992 survey, however, a special procedure was used to impute literacy proficiencies for literacy-related nonrespondents.

Due to such procedural differences, direct comparisons of the results of the two surveys are not simple and straight-forward. However, because the 1992 sample is more inclusive than the 1985 sample, subsamples that have more exact counterparts in the 1985 survey can be selected. For instance, the initial report from the 1992 NALS presented data, using no subsample matching that indicated that young adults in 1992 were somewhat less literate than their predecessors in 1985. However, when a comparison was made between matched subsamples of the 1985 and 1992 survey respondents based on reasons for nonresponse, the proficiency differences decreased significantly. Furthermore, results from partition analysis of the two surveys' matched subsamples—based on change due to variations in demographic characteristics versus change not related to demography—suggest that most of the observed declines in the average literacy skills of young adults over time can be accounted for by shifts in the composition of the population and by changes across the assessments in the rules used to include or exclude nonrespondents.

Comparisons with the 1993 General Educational Development (GED) Tests. Comparisons between NALS and GED examinees are explored in *The Literacy Proficiencies of GED Examinees: Results From the GED-NALS Comparison Study* (Baldwin et al. 1993). The GED tests and NALS instruments have a considerable degree of overlap in what they measure. Both assess skills that appear to represent verbal comprehension and reasoning or the ability to understand, analyze, interpret, and evaluate written information and apply fundamental principles and concepts. Despite the considerable degree of overlap, the two instruments also measure somewhat different skills. For example, the GED tests seem to tap unique dimensions of writing mechanics and mathematics, while the adult literacy scales appear to tap unique dimensions of document literacy. In addition, the evidence shows that there are no differences in the average prose, document, or quantitative literacy skills of those adults who terminated their schooling at the high school or GED level.

6. CONTACT INFORMATION

For content information on the National Adult Assessments of Literacy, contact

Andrew J. Kolstad
Phone: (202) 502-7374
E-mail: andrew.kolstad@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences

U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Baldwin, J., Kirsch, I.S., Rock, D., and Yamamoto, K. (1993). *The Literacy Proficiencies of GED Examinees: Results From the GED-NALS Comparison Study*. American Council on Education and Educational Testing Service.

Kirsch, I.S., Jungeblut, A., Jenkins, L., and Kolstad, K. (1993). *Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey* (NCES 93-275). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Kirsch, I.S., Yamamoto, K., Norris, N., Rock, D., Jungeblut, A., O'Reilly, P., Campbell, A., Jenkins, L., Kolstad, A., Berlin, M., Mohadjer, L., Waksberg, J., Goksel, H., Burke, J., Rieger, S., Green, J., Klein, M., Mosenthal, P., and Baldi, S. (2000). *Technical Report and Data File User's Manual for the 1992 National Adult Literacy Survey* (NCES 2001-457). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Survey Design

Campbell, A., and Kirsch, I.S. (1992). *Assessing Literacy: The Framework for the National Adult Literacy Survey* (NCES 92-113). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Chapter 20: National Assessment of Adult Literacy (NAAL)

1. OVERVIEW

The 2003 National Assessment of Adult Literacy (NAAL) is a nationally representative assessment of English literacy among American adults age 16 and older. Sponsored by the National Center for Education Statistics (NCES), NAAL is the nation's most comprehensive measure of adult literacy since the 1992 National Adult Literacy Survey (NALS).

In 2003, over 19,000 adults participated in the national and state-level assessments, representing the entire population of U.S. adults age 16 and older (in households and prisons) in the 50 states and the District of Columbia. Approximately 1,200 of the participants were inmates of state and federal prisons who were assessed separately in order to provide estimates of literacy for the incarcerated population.

By comparing results from 1992 and 2003, NAAL provides the first indicator in a decade of the nation's progress in adult literacy. NAAL also provides information on adults' literacy performance and related background characteristics to researchers, practitioners, policymakers, and the general public.

Purpose

To (1) evaluate the English language literacy skills of adults (age 16 and older) living in households or prisons in the United States; (2) relate the literacy skills of the nation's adults to a variety of demographic characteristics and explanatory variables; and (3) compare the results with those from the 1992 NALS.

Components

NAAL includes a number of components that capture the breadth of adult literacy in the United States: the Background Questionnaire helps identify the relationships between adult literacy and selected demographic and background characteristics; the Prison Component assesses the literacy skills of adults in federal and state prisons; the State Assessment of Adult Literacy (SAAL) gives statewide estimates of literacy for states participating in the state-level assessment; the Health Literacy Component introduces the first-ever national assessment of adults' ability to use their literacy skills in understanding health-related materials and forms; the Fluency Addition to NAAL (FAN) measures basic reading skills by assessing adults' ability to decode, recognize words, and read with fluency; the Adult Literacy Supplemental Assessment (ALSA) provides information on the ability of the least literate adults to identify letters and numbers and to comprehend simple prose and documents; and the main assessment offers a picture of the general literary (i.e., prose, document and quantitative literary) of the adults who passed the core literary tasks.

SURVEY OF A SAMPLE OF ADULTS LIVING IN HOUSEHOLDS OR PRISONS:

Assesses literacy skills:

- Prose
- Document
- Quantitative

Collects background data on:

- Demographics
- Education
- Labor Market Experiences
- Income
- Activities

Background Questionnaire. The 2003 NAAL Background Questionnaire collected data in a variety of background categories; it obtained valuable background information not collected in the 1992 survey. The questionnaire served three purposes:

- to provide descriptive data on respondents;
- to enhance understanding of the factors that are associated with literacy skills used at home, at work, or in the community; and
- to allow for the reporting of changes over time.

The questionnaire was orally administered to every participant by an interviewer who used a computer-assisted personal interview (CAPI) system. Unlike the 1992 NALS, in which the background questions were read aloud from a printed questionnaire, in 2003, interviewers read the questions from laptop computer screens and entered the responses directly into the computer. CAPI then selected the next question based on responses to prior questions. Because the questions were targeted, a respondent did not answer all of the background questions (i.e., inapplicable questions were skipped). The questionnaire took about 28 minutes to complete.

The background questionnaire used in SAAL was the same as that used in NAAL. However, a separate questionnaire was administered for the prison component in order to address issues of particular relevance to the prison population.

Prison Component. The 2003 NAAL Prison component assesses the literacy skills and proficiencies of the U.S. adult prison population. In the 2003 assessment, approximately 1,200 adults participated, from 107 prisons (including 12 federal prisons) in 31 states.

Key features:

- provides demographic and performance data for the prison population, in comparison with the main NAAL household study of the general adult population;
- reports results that are useful to policymakers and practitioners concerned with literacy and education in correctional settings; and
- guides corrections and education professionals in the development of more effective literacy and adult education programs for prison inmates.

The principal aim of the 2003 NAAL prison component is to provide comprehensive information on the literacy and background of the U.S. adult prison population to policymakers and practitioners in order to enhance adult education in our nation's prisons and improve incarcerated adults' ability to function and achieve their goals in the general society, in the workplace, at home, and in the community–upon their release from prison.

State Assessment of Adult Literacy (SAAL). The SAAL is an assessment of adult literacy within a participating state. Conducted in conjunction with the 2003 NAAL data collection, SAAL collected additional data within the six participating states: Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma.

Key features:

SAAL provides participating states with individually-tailored reports that offer:

- more in-depth analysis of a state's literacy, by augmenting the state's sample with the national sample;
- state and national comparisons;
- expanded background information on population groups;
- state-level scoring for FAN, ALSA, and the Health Literacy Component;
- estimates by demographic and other characteristics of interest; and
- trend data (for New York), because it participated in both the 1992 and 2003 assessments.

Health Literacy Component. The 2003 NAAL is the first large-scale national assessment in the United States to contain a component designed specifically to measure health literacy—the ability to use literacy skills to read and understand written health-related information encountered in everyday life. The Health Literacy Component establishes a baseline against which to measure progress in health literacy in future assessments.

The NAAL health literacy report—*The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy* (Kutner et al. 2007)—provides first-hand information on the status

of the health literacy of American adults age 16 and older. Results are reported in terms of the four literacy performance levels—below basic, basic, intermediate, and proficient—with examples of the types of health literacy tasks that adults at each level may be able to perform.

Key features:

- reports on the health literacy skills of target audiences;
- sheds light on the relationship between health literacy and background variables, such as educational attainment, age, race/ethnicity, adults' sources of information about health issues, and health insurance coverage;
- examines how health literacy is related to prose, document, and quantitative literacy;
- provides information that may be useful in the development of effective policies and customized programs that address deficiencies in health literacy skills; and
- guides the development of health information tailored to the strengths and weaknesses of target audiences.

Fluency Addition to NAAL (FAN). FAN examines components of oral reading fluency that the main NAAL does not assess. Using speech-recognition software, FAN measures adults' ability to decode, recognize words, and read with fluency.

Key features:

- establishes a basic reading skills scale;
- identifies, for the first time, the relationship between basic reading skills and selected background characteristics, as well as performance on the main NAAL, Health Literacy Component, and prison component; and
- provides a baseline for measuring future changes in the levels and distribution of oral fluency over time.

Ultimately, FAN can improve our understanding of the skill differences between adults who are able to perform relatively challenging tasks and adults who lack basic reading skills. Such information will prove most useful to researchers, practitioners, and policymakers. For instance, adult education providers can use FAN results to develop and offer instruction and courseware that will better address the skill sets of the least literate adults. Likewise, policymakers can use FAN results to support the creation and improvement of programs serving adults with lower literacy skills.

Adult Literacy Supplemental Assessment (ALSA). Low levels of literacy are likely to limit life chances and may be related to social welfare issues, including poverty, incarceration, and preventive health care. Given this, it has become increasingly important for researchers, policymakers, and practitioners to understand the literacy skills and deficits of the least literate adults.

ALSA is designed to assess the basic reading skills of the least literate adults. The 1992 NALS lacked a similar component. Because the least literate adults were unable to complete the 1992 assessment due to literacy-related complications (e.g., difficulty reading and writing in English; mental or learning disabilities), the 1992 NALS provided little information on these respondents.

Key features:

- enhances our understanding of the basic reading skills of the least-literate adults;
- identifies relationships between ALSA scores and selected background characteristics of adults;
- reports results for appropriate demographic groups (e.g., Black, Hispanic, and other racial/ethnic groups; ESL adults; the prison population);
- describes relationships between the performance of ALSA participants and main NAAL participants on the FAN oral reading tasks; and
- provides a baseline for measuring future changes in the levels and distribution of the least literate adults' basic reading skills over time.

Participants who scored low on the core screening questions (see "Assessment Design" below) were given ALSA instead of the main assessment.

The Main Assessment. NAAL main assessment reports a separate score for each of three literacy

areas: prose literacy, document literacy, and quantitative literacy.

Prose literacy refers to the knowledge and skills needed to perform prose tasks—that is, to search, comprehend, and use continuous texts. Prose examples include editorials, news stories, brochures, and instructional materials.

Document literacy refers to knowledge and skills needed to perform document tasks—that is, to search, comprehend, and use continuous texts. Document examples include job applications, payroll forms, transportation schedules, maps, tables, and drug or food labels.

Quantitative literacy refers to the knowledge and skills needed to perform prose tasks—that are, to identify and perform computations, either alone or sequentially, using numbers embedded in printed materials. Examples include balancing a check book, computing a tip, completing and order form, or determining the amount of interest on a loan from an advertisement.

Periodicity
The 2003 NAAL results are comparable to those of the 1992 NALS, and for young adults 21 to 25 years old, to the 1985 young adult literacy assessment.

2. USES OF DATA

NAAL data provide vital information to policymakers, business and labor leaders, researchers, and citizens. The survey results can be used to

- describe the levels of literacy demonstrated by the adult population as a whole and by adults in various subgroups (e.g., those targeted as at risk, prison inmates, and older adults);
- characterize adults' literacy skills in terms of demographic and background information (e.g., reading characteristics, education, and employment experiences);
- profile the literacy skills of the nation's workforce;
- compare assessment results from the current study with those from the 1992 NALS;
- interpret the findings in light of information-processing skills and strategies, so as to inform curriculum decisions concerning adult education and training; and
- increase our understanding of the skills and knowledge associated with living in a technological society.

3. KEY CONCEPTS

NAAL is designed to measure functional English literacy. The assessment measures how adults use printed and written information to adequately function at home, in the workplace, and in the community.

Since adults use different kinds of printed and written materials in their daily lives, NAAL measures three types of literacy—prose, document, and quantitative—and reports a separate scale score for each of these three areas. By measuring literacy along three scales, instead of just one, NAAL can provide more comprehensive data on literacy tasks and literacy skills associated with the broad range of printed and written materials adults use.

Prose Literacy
The prose literacy scale measures the knowledge and skills needed to perform prose tasks (i.e., to search, comprehend, and use continuous texts). Examples include editorials, news stories, brochures, and instructional materials.

Document Literacy
The document literacy scale measures the knowledge and skills needed to perform document tasks (i.e., to search, comprehend, and use non-continuous texts in various formats). Examples include job applications, payroll forms, transportation schedules, maps, tables, and drug or food labels.

Quantitative Literacy
The quantitative literacy scale measures the knowledge and skills required to perform quantitative tasks (i.e., to identify and perform computations, either alone or sequentially, using numbers embedded in printed materials). Examples include balancing a checkbook, figuring out a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

In addition to the prose, document, and quantitative literacy scales, the 2003 assessment included a health literacy scale. The health literacy scale contains prose, document, and quantitative items with health-related

content. The items fall into three areas: clinical, prevention, and navigation of the health system.

4. SURVEY DESIGN

Data collection for the main NAAL study and the concurrent state assessment, SAAL, was conducted in 2003 using in-person household interviews. Over 18,000 adults participated, selected from a sample of over 35,000 households that represented the entire U.S. household population age 16 and over—about 222 million Americans (U.S. Census Bureau, Current Population Survey 2003). In addition, approximately 1,200 inmates from 110 federal and state prisons were assessed in early 2004 for the prison component, which provides separate estimates of literacy levels for the incarcerated population.

All household participants received an incentive payment of $30 in an effort to increase both the representativeness of the sample and the response rate. Black and Hispanic households were oversampled at the national level to ensure reliable estimates of their literacy proficiencies. Special accommodations were made for adults with disabilities or with limited English proficiency.

Target Population

The target population for the national household sample consisted of adults 16 and older in the 50 states and the District of Columbia who, at the time of the survey, resided in private households or college dormitories. The target population for the supplemental state household sample consisted of individuals 16 to 64 years of age who, at the time of the survey, resided in private households or college dormitories in the participating state. The target population for the prison sample consisted of inmates age 16 and older in state and federal prisons at the time of the survey; those held in local jails, community-based facilities, or other types of institutions were not included.

Sample Design

The 2003 NAAL included two samples: (1) adults age 16 and older living in households (99 percent of the entire NAAL sample, weighted); and (2) inmates age 16 and older in state and federal prisons (1 percent of the entire NAAL sample, weighted). Each sample was weighted to represent its share of the total population of the United States, and the samples were combined for reporting.

Household sample. The 2003 NAAL household sample included a nationally representative probability sample of 35,000 households. The household sample was selected on the basis of a four-stage, stratified area sample: (1) primary sampling units (PSUs) consisting of counties or groups of contiguous counties; (2) secondary sampling units (referred to as segments) consisting of area blocks; (3) housing units containing households; and (4) eligible persons within households. Person-level data were collected through a screener, a background questionnaire, the literacy assessment, and the oral module.

Six states—Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma—purchased additional cases in their states to allow reporting at the state level. A single area sample was selected for the national NAAL sample, and additional samples were selected for the six states participating in the SAAL. For each sample, the usual procedures for area sampling were followed: a stratified probability proportionate to size design was used for the first two stages, and systematic random samples were drawn in the last two stages.

A key feature of the national NAAL sample was the oversampling of Black and Hispanic adults, which was accomplished by oversampling segments with high concentrations of these groups. The SAAL samples did not include any oversampling of Black, Hispanic, or other racial/ethnic groups.

Although integrating the NAAL and SAAL samples at the design stage would have been more effective statistically, the states agreed to participate after the NAAL sample design and selection process had been finalized. Therefore, the approach used in the 1992 NALS was followed: selecting the SAAL samples independently of the NAAL sample and combining the samples at the estimation phase by using composite estimation.

Stage one sampling. The first stage of sampling was the selection of PSUs, which consisted of counties or groups of counties. PSUs were formed within state boundaries, which gave an improved sample for state-level estimation. One PSU was selected per stratum by using probabilities proportionate to their population within households, except in Maryland and Massachusetts, where samples of segments were selected as the first-stage units. One hundred PSUs were selected for the national sample, and 54 PSUs were selected in Kentucky, Missouri, New York, and Oklahoma. Maryland and Massachusetts had too few PSUs from which to sample; therefore, segments were selected in the first stage of sampling. After selecting the segments, 20 area clusters (quasi-PSUs) were

created for Maryland and Massachusetts by grouping the selected segments into 20 geographically clustered areas to facilitate a cost-efficient approach to data collection. The true first-stage sample size is much larger because a total of 323 first-stage units (i.e., segments) were selected in Maryland and Massachusetts. Fourteen PSUs were selected for both the national NAAL and the SAAL samples; hence, the sample included a combined total of 160 unique PSUs.

Stage two sampling. In the second stage of sampling, segments (census blocks or groups of blocks) within the PSUs were selected with a probability proportionate to size; the measure of size for a segment was a function of the number of year-round housing units within the segment. In the national sample, the Black and Hispanic populations were sampled at a higher rate than the remainder of the population to increase their sample size, whereas the state samples used no oversampling. Oversampling in the national sample was accomplished by oversampling the segments in which Black and Hispanic adults accounted for 25 percent or more of the population. There were 2,000 segments selected for the national sample and 861 segments selected across the SAAL samples, with a total of 2,800 unique segments selected across the national and six SAAL samples. (Two segments were selected for both the NAAL and SAAL samples.)

Stage three sampling. In the third stage of sampling, housing units were selected with equal probability within each segment, except for White households within high percentage of Black, Hispanic, and other race/ethnicity segments in the national component. These national sample households were subsampled after screening so that the sampling rates for White persons would be about the same in the high percentage of Black, Hispanic, and other race/ethnicity segments as in other segments. The overall sample size of housing units took into account expected losses owing to vacant housing units, units that were not housing units, and expected response rates.

Stage four sampling. The fourth stage of selection involved listing the age-eligible household members (age 16 and older) for each selected household. Subsequently, one person was selected at random within households with three or fewer eligible persons, and two persons were selected if the household had four or more eligible persons. The listing and selection of persons within households were performed with the CAPI system.

Of the 35,000 sampled households, 4,700 were either vacant or not a dwelling unit, resulting in a sample of 31,000 households. A total of 25,000 households completed the screener, which was used to select survey respondents. The final screener response rate was 81 percent (weighted).

On the basis of the screener data, 24,000 respondents age 16 and older were selected to complete the background questionnaire and the assessment; 18,000 actually completed the background questionnaire. Of the 5,500 respondents who did not complete the background questionnaire, 360 were unable to do so because of a literacy-related barrier, either the inability to communicate in English or Spanish (the two languages in which the background questionnaire was administered) or a mental disability.

The final response rate for the background questionnaire—which included respondents who completed the background questionnaire and respondents who were unable to complete the background questionnaire because of language problems or a mental disability—was 77 percent (weighted). Of the 18,000 adults age 16 and older who completed the background questionnaire, 17,000 completed at least one question on each of the three scales—prose, document, and quantitative—measured in the adult literacy assessment. An additional 149 were unable to answer at least one question on each of the three scales for literacy-related reasons or a mental disability. The final response rate for the literacy assessment—which included respondents who answered at least one question on each scale plus the 150 respondents who were unable to do so because of language problems or a mental disability—was 97 percent (weighted).

Cases were considered complete if the respondent completed the background questionnaire and at least one question on each of the three scales or if the respondent was unable to answer any questions because of language issues (an inability to communicate in English or Spanish) or a mental disability. All other cases that did not include a complete screener, a background questionnaire, and responses to at least one question on each of the three literacy scales were considered incomplete or missing. Before imputation, the overall response rate for the household sample was 60 percent (weighted).

Imputation for nonresponse. For respondents who did not complete any literacy tasks on any scale, no information is available about their performance. Completely omitting these individuals from the analyses would have resulted in unknown biases in

estimates of the literacy skills of the national population because refusals cannot be assumed to have occurred randomly. For 860 respondents who answered the background questionnaire but refused to complete the assessment for reasons other than language issues or a mental disability, regression-based imputation procedures were applied to impute responses to one assessment item on each scale by using the NAAL background data on age, gender, race/ethnicity, education level, country of birth, census region, and metropolitan statistical area status.

On the prose and quantitative scales, a response was imputed for the easiest task on each scale. On the document scale, a response was imputed for the second easiest task because that task was also included on the health literacy scale. In each of the logistic regression models, the estimated regression coefficients were used to predict missing values of the item to be imputed. For each nonrespondent, the probability of answering the item correctly was computed and then compared with a randomly generated number between 0 and 1. If the probability of getting a correct answer was greater than the random number, the imputed value for the item was 1 (correct); otherwise, it was 0 (wrong). In addition, a wrong response on each scale was imputed for 65 respondents who started to answer the assessment, but were unable to answer at least one question on each scale because of language issues or a mental disability.

The final household reporting sample—including the imputed cases—consisted of 18,000 respondents. These 18,000 respondents include the 17,000 respondents who completed the background questionnaire and the assessment; the 860 respondents who completed the background questionnaire, but refused to do the assessment for non-literacy-related reasons (and have imputed responses to one item on each scale); and the 70 respondents who started to answer the assessment items, but were unable to answer at least one question on each scale because of language issues or a mental disability. After including the cases for which responses to the assessment questions were imputed, the weighted response rate for the household sample was 62 percent (18,000 cases with complete or imputed data and an additional 440 cases that had no assessment data because of language issues or a mental disability).

Prison sample. The 2003 assessment also included a nationally representative probability sample of inmates in state and federal prisons. The target population for the prison sample consisted of inmates age 16 and older from state and federal prisons in the United States. The sampling frame was created primarily from two data sources: the Bureau of Justice Statistics 2000 Census of State and Federal Adult Correctional Facilities (referred to in the following text as the Prison Census) and the 2003 Directory of Correctional Facilities of the American Correctional Association (ACA).

The facility universe for the NAAL Prison Component was consistent with the Prison Census. As defined for the Prison Census, the 2003 NAAL target population included the following types of state and federal adult correctional facilities: prisons; prison farms; reception, diagnostic, and classification centers; road camps; forestry and conservation camps; youthful offender facilities (except in California); vocational training facilities; drug and alcohol treatment facilities; and state-operated local detention facilities in Alaska, Connecticut, Delaware, Hawaii, Rhode Island, and Vermont. Facilities were included in the NAAL Prison Component if they were:

- staffed with federal, state, local, or private employees;
- designed to house primarily state or federal prisoners;
- physically, functionally, and administratively separate from other facilities; and
- in operation between September 2003 and March 2004.

Specifically excluded from the NAAL Prison Component were:

- privately operated facilities that were not exclusively for state or federal inmates;
- military facilities;
- Immigration and Naturalization Service facilities;
- Bureau of Indian Affairs facilities;
- facilities operated and administered by local governments, including those housing state prisoners;
- facilities operated by the U.S. Marshals Service, including the Office of the Detention Trustee;

- hospital wings and wards reserved for state prisoners; and
- facilities housing only juvenile offenders.

Even though they contain inmates up to age 21, juvenile facilities were excluded from NAAL for two reasons: (1) to remain consistent with the facilities listed in the Prison Census; and (2) to promote cost efficiency because it would not have been cost-effective to visit these facilities to sample the small number of inmates 16 years of age and older.

Inmate sampling frames were created by interviewers at the time they visited the prisons. The frame consisted of all inmates occupying a bed the night before inmate sampling was conducted.

Approximately 110 prisons were selected to participate in the adult literacy assessment. The final prison response rate was 97 percent (weighted). Among the inmates in these prisons, 1,300 inmates ages 16 and older were randomly selected to complete the background questionnaire and assessment. Of these 1,300 selected inmates, 1,200 completed the background questionnaire. Of the 140 inmates who did not complete the background questionnaire, about 10 were unable to do so because of a literacy-related barrier (either the inability to communicate in English or Spanish) or a mental disability.

The final response rate for the prison background questionnaire—which included respondents who completed the background questionnaire and respondents who were unable to complete the background questionnaire because of language problems or a mental disability—was 91 percent (weighted). Of the 1,200 inmates who completed the background questionnaire, 1,100 completed at least one question on each of the three scales—prose, document, and quantitative—measured in the adult literacy assessment. An additional 10 inmates were unable to answer at least one question on each of the three scales for literacy-related reasons. The final response rate for the literacy assessment—which included respondents who answered at least one question on each scale or were unable to do so because of language problems or a mental disability—was 99 percent (weighted).

The same definition of a complete case used for the household sample was also used for the prison sample, and the same rules were followed for imputation. Before imputation, the final response rate for the prison sample was 87 percent (weighted).

Imputation for nonresponse. One response on each scale was imputed on the basis of background characteristics for 30 inmates who completed the background questionnaire, but had incomplete or missing assessments for reasons that were not literacy related. The statistical imputation procedures were the same as for the household sample. The background characteristics used for the missing data imputation for the prison sample were prison security level, region of country/type of prison, age, gender, educational attainment, country of birth, race/ethnicity, and marital status. A wrong response on each scale was imputed for the inmates who started to answer the assessment, but were unable to answer at least one question on each scale because of language issues or a mental disability. The final prison reporting sample—including the imputed cases—consisted of 1,200 respondents. After the cases for which responses to the assessment questions were imputed were included, the weighted response rate for the prison sample was 88 percent (1,200 cases with complete or imputed data and an additional 20 cases that had no assessment data because of language issues or a mental disability).

Assessment Design

The NAAL interview was conducted in the order described below.

First, every respondent completed a background questionnaire that collected data on demographic, socioeconomic, and other factors associated with literacy.

Next, every respondent completed seven core screening questions, which were among the easiest in the assessment.

Similar in structure to the main NAAL assessment questions, the core questions determined whether a respondent's skills were sufficient to participate in the main NAAL assessment or if the individual should be routed to ALSA. Interviewers used a scoring rubric to code respondents' answers to each code question (e.g., "1" for correct, "2" for wrong, and "3" for no response). Interviewers entered the codes into a CAPI System, which selected respondents for ALSA using an empirically derived algorithm that predicts very low performance on the main NAAL. ALSA assessed the ability of the least literate adults to identify letters and numbers and to comprehend simple prose materials. Those participants who scored low on the basic core screening questions took ALSA instead of the main NAAL.

After completing either the main NAAL assessment booklet or ALSA, every respondent took FAN. FAN used speech-recognition software to assess adults' ability to decode and recognize words and to read with fluency.

Data Collection and Processing

Reference dates. Household data collection was conducted from March 2003 through February 2004; prison data collection was conducted from March through July 2004.

Data collection. Household interviews took place in respondents' homes; prison interviews generally took place in a classroom or library in the prison. Whenever possible, interviewers administered the background questionnaire and assessment in a private setting. Unless there were security concerns, a guard was not present in the room when inmates were interviewed.

Interviewers used a CAPI system programmed into laptop computers. The interviewers read the background questions from the computer screen and entered all responses directly into the computer. Skip patterns and follow-up probes for contradictory or out-of-range responses were programmed into the computer.

After completing the background questionnaire, respondents were handed a booklet with the assessment questions. The interviewers followed a script that introduced the assessment booklet and guided the respondent through the assessment.

Each assessment booklet began with the same seven screening questions. After the respondent completed the screening questions, the interviewer asked the respondent for the book and used an algorithm to determine, on the basis of the responses to the questions, whether the respondent should continue in the main assessment or be placed in ALSA. Three percent (weighted) and 5 percent (unweighted) of adults were placed in the ALSA.

ALSA is a performance-based assessment that allowed adults with marginal literacy to demonstrate what they could and could not do when asked to make sense of various forms of print. The ALSA started with simple identification tasks and sight words and moved to connected text, using authentic, highly contextualized material commonly found at home or in the community.

Respondents were routed to an alternative assessment (ALSA) based on their performance on the seven easy screening tasks at the beginning of the literacy assessment. Because the ALSA respondents answered most, or all, of these questions incorrectly, if they were place on the NAAL scale, they would have been classified on the NAAL scale as below basic level on the health scale.

A respondent who continued in the main assessment was given back the assessment booklet, and the interviewer asked the respondent to complete the tasks in the booklet and guided the respondent through them. The main assessment consisted of 12 blocks of tasks with approximately 11 questions in each block, but each assessment booklet included only 3 blocks of questions. The blocks were spiraled so that across the 26 different configurations of the assessment booklet, each block was paired with every other block and each block appeared in each of the three positions (first, middle, last) in a booklet.

For ALSA interviews, the interviewer read the ALSA script from a printed booklet and classified the respondent's answers into the response categories in the printed booklet. ALSA respondents were handed the materials they were asked to read.

Following the main assessment or ALSA, all respondents were administered FAN (the oral fluency assessment). Respondents were handed a booklet with passages, number lists, letter lists, word lists, and pseudoword lists to read orally. Respondents read into a microphone that recorded their responses on the laptop computer.

Accommodations. With the passage of the Americans with Disabilities Act and the growth of America's immigrant population, assessment programs like NAAL must consider issues of inclusion and accommodation. The 2003 NAAL provided for two types of accommodations—administrative and language.

Administrative accommodations were made for adults with disabilities. First, NAAL is inherently accommodating because the assessment was conducted one-on-one in the respondent's home. Second, all respondents with disabilities received additional time to complete the assessment, if necessary.

Language accommodations were made for adults with limited English proficiency or whose primary language is not English. Questions on the background questionnaire were available in either English or Spanish. In addition, instructions for FAN, ALSA, and the core screening test questions were given in either English or Spanish. However, the stimulus materials

for these questions were in English since NAAL's main objective is to assess literacy in English.

Results are reported separately for non-native speakers of English and compared to the results of native speakers of English. Thus, the unique needs of English as a Second Language (ESL) adults may be better understood by researchers, policymakers, and practitioners.

Data processing. The NAAL assessment questions were open-ended and thus required scoring by trained scorers. NAAL experts have developed scoring rubrics that detail the rules necessary for scoring each assessment question.

In order to make NAAL scores meaningful, the scores were grouped into performance levels to provide information that could more easily be understood and used by the public and policymakers. The performance levels were developed to characterize the status of English language literacy of American adults and include the following: nonliterate in English, below basic, basic, intermediate, and proficient literacy. For reporting purposes adults classified as nonliterate in English are included in the below basic literacy level. The 2003 NAAL performance levels are different from the five levels NCES used to report NALS results in 1992. However, in order to make comparisons across years, the 1992 data were reanalyzed and the new performance levels were applied to the 1992 data.

NAAL scoring is designed to measure adults' abilities to perform literacy tasks in everyday life. Since adults are likely to make mistakes as they interact with printed and written material, NAAL scorers make allowances for partial responses and writing errors.

While most responses are either correct or incorrect, a response can be partially correct if the information provided is still useful in accomplishing the task. For example, a respondent who writes the wrong product price on a catalog order form could receive partial credit, because in real life such a minor error would not necessarily result in the placement of an incorrect order (since other information is provided, such as product name and price). However, if a respondent miswrites a social security number on a government application form, such an error would not receive partial scoring.

Similarly, responses containing writing errors—grammatical and spelling errors, use of synonyms, incomplete sentences, or circling instead of writing the correct answer—are scored as correct as long as the overall meaning is correct and the information provided accomplishes the task. However, if a respondent is filling out a form and writes the answer on the wrong line, or if, for a quantitative task, the calculation is right but the respondent writes the wrong answer in the blank, then the response is scored as incorrect.

During the task development stage, scoring experts developed scoring rubrics that detailed the rules for scoring each assessment question. To ensure that all assessment questions were scored accurately, NAAL scoring rubrics underwent several stages of verification both before and after the assessment was administered.

Before the main NAAL study began, a field test of about 1,400 adults was conducted to help identify and screen out problems with the scoring rubrics, such as alternative correct responses and scoring rubrics that are difficult to implement consistently (thus leading to low rates of interrater reliability).

After the main study ended, a sample of responses from the household and prison interviews was scored using the scoring rubrics. As the test developers scored the sample responses, they made adjustments to the scoring rubrics to reflect the kinds of responses adults gave during the assessment. Together, these sample responses and the revised scoring rubrics were used in training the scorers who scored the entire assessment.

In a group setting, scorers were trained to recognize each task and its corresponding scoring rubric, as well as sample responses that are representative of correct, partially correct, and incorrect answers. After group training, readers scored numerous practice questions before they began to score actual booklets.

To ensure that readers were scoring accurately, 50 percent of the assessment questions were subject to a second interrater reliability check, in which a second reader scored the booklet and the scores of the first and second readers were compared. Interrater reliability is the percentage of times two readers agree exactly in their scores. (In 1992, the average percentage of agreement was 97 percent.) Any batch of questions that exceeded a low level of scoring mistakes was sent back to the scorers for corrections. Also, the scoring supervisor discussed the discrepancy with the scorers involved. Quality control procedures like this ensured reliability of the scoring.

Performance levels. Performance levels are important because they provide the ability to group people with similar literacy scores into a relatively small number of categories of importance to the adult education community, much like grouping students with similar scores on a test into various letter grades (e.g., A or B).

A benefit of having performance levels is that they enable NAAL to characterize American adults' relative literacy strengths and weaknesses by describing the nature and difficulty of the literacy tasks that participants at each level can perform with a reasonably high rate of success.
Performance levels were determined in response to a request from NCES to the National Research Council (NRC), which convened a Committee on Performance Levels for Adult Literacy. The committee's goal was to do the following in an open and public way: evaluate the literacy levels used by NAAL's 1992 predecessor survey, and recommend a set of performance levels that could be used in reporting the 2003 results and also be applied to the 1992 results in order to make comparisons across years.

New performance levels. After reviewing information about the 1992 and 2003 assessments as well as feedback from stakeholders (e.g., adult literacy practitioners), the NRC committee specified a new set of performance levels intended to correspond to four policy-relevant categories of adults, including adults in need of basic adult literacy services. The next step was to determine the score ranges to be included in each level for each of the three NAAL literacy scales—prose, document, and quantitative literacy.

Score ranges. To determine the score ranges for each level, the committee decided to use the "bookmark" method. Initial implementation of the method involved describing the literacy skills of adults in the four policy-relevant levels, and holding two sessions with separate panels of "judges" consisting of adult literacy practitioners, officials with state offices of adult education, and others. One group of judges focused on the 1992 assessment tasks and the other group focused on the 2003 assessment tasks.

Bookmarks. For each literacy area (prose, document, and quantitative), the judges were given, in addition to descriptions of the performance levels, a booklet of assessment tasks arranged from easiest to hardest. The judges' job was to place "bookmarks" in the set of tasks that adults at each level were "likely" to get right. The term "likely" was defined as "67 percent of the time," or two out of three times, and statistical procedures were used to determine the score associated with a 67 percent probability of performing the task correctly. The bookmarks designated by the judges at the two sessions were combined to produce a single bookmark-based cut score for each performance level on each of the three literacy scales.

Quasi-contrasting groups approach. To refine the bookmark-based cut scores, which indicated the lowest score to be included in each performance level, the committee used a procedure it termed the "quasi-contrasting groups approach." The committee compared the 2003 bookmark-based cut scores with the 1992 scores associated with various background variables, such as educational attainment. The criterion for selecting the background variables was potential usefulness for distinguishing between adjacent performance levels, such as basic and below basic (e.g., having some high school education vs. none at all; reporting that one reads well vs. not well; reading a newspaper sometimes vs. never reading a newspaper; reading at work sometimes or more often vs. never reading at work).

In each case, the midpoint between the average scores of the two adjacent performance levels (below basic and basic; basic and intermediate; intermediate and proficient) was calculated and averaged across the variables that provided contrast between the groups. The committee developed a set of rules and procedures for deciding when and how to make adjustments to the bookmark cut scores when the cut scores associated with the selected background variables were different from the bookmark-based scores.

Nonliterate in English classification. The NRC committee recommended that NCES distinguish a fifth group of adults with special importance to literacy policy—those who are nonliterate in English. As originally defined by the committee, this category consisted of adults who performed poorly on a set of easy screening tasks in 2003 and therefore were routed to an alternative assessment for the least literate adults (i.e., ALSA). Because the 1992 assessment included neither the alternative assessment nor the 2003 screening tasks, adults in this category cannot be identified for 1992.

To provide a more complete representation of the adult population that is nonliterate in English, NCES expanded the category to include not only the 3 percent of adults who took the alternative assessment, but also the 2 percent who were unable to be tested at all because they knew neither English nor Spanish (the other language spoken by interviewers). Thus, as defined by NCES, the category included about 5 percent of adults in 2003.

Refinements made before using the new levels. The new performance levels were presented to NCES as recommendations. Having accepted the general recommendations, NCES incorporated a few refinements before using the levels to report results. First, NCES changed the label of the top category from advanced to proficient because the term "proficient"

better conveys how well the upper category of adults performs. Second, NCES added sample tasks from the 2003 assessment to illustrate the full range of tasks that adults at each level can perform, as well as a brief (one-sentence) summary description for each level to enhance public understanding. Third, as outlined in the previous paragraph, NCES included additional adults in the “nonliterate in English” category.

Estimation Methods

Weighting. As discussed above, NAAL included both a household sample and a prison sample. The household sample was further divided into the cases selected for the national sample and the additional cases selected in the six SAAL states. Weighting was done separately for the household and prison samples. However, the weights were developed so that the two samples could be used together in a combined sample.

Household sample weighting. Differential probabilities of selection into the NAAL household sample were adjusted by computing base weights for all adults selected into the sample. The base weight was calculated as the reciprocal of a respondent’s final probability of selection. The weights were adjusted for nonresponse at both the screener level and the background questionnaire level. Additionally, trimming procedures were followed to reduce the impact of extreme weights. The background questionnaire weighting steps were done separately for the national and SAAL household samples, and each sample was calibrated separately to population estimates based on 2003 Current Population Survey (CPS) data. To combine the NAAL and SAAL household samples, composite weights were calculated for the respondents in the six participating SAAL states and the respondents in the national NAAL household sample in these six states. The composite weights were adjusted through poststratification and raking to match the 2003 CPS data.

Prison sample weighting. The prison component weighting consisted of four main steps. First, prison base weights were constructed using the probability of selection for each prison into the sample. Then, a nonresponse adjustment was made to the prison base weights to account for nonparticipating prisons. Next, inmate base weights were calculated using the prison nonresponse-adjusted weight and the within-prison sampling rate. Finally, the inmate base weights were raked to Bureau of Justice Statistics control totals to account for inmate nonresponse and noncoverage.

Variance estimation. A complex sample design was used to select assessment respondents. The properties of a sample selected through a complex design can be very different from those of a simple random sample. (In a simple random sample, every individual in the target population has an equal chance of selection and the observations from different sampled individuals can be considered to be statistically independent of one another.) Sampling weights should be used to account for the fact that the probabilities of selection were not identical for all respondents. All population and subpopulation characteristics based on the NAAL data should use sampling weights in their estimation.

Since the respondents were selected using complex sample design, conventional formulas for estimating sampling variability that assume simple random sampling (and, hence, independence of observations) are inappropriate. Standard errors calculated as though the data had been collected from a simple random sample would generally underestimate sampling errors. Therefore, the properties of the complex data collection design should be taken into account during the analysis of the data.

Scaling. Each respondent to NAAL received a booklet that included 3 of the 13 assessments blocks. Because each respondent did not answer all of the NAAL items, item response theory (IRT) methods were used to estimate average scores on the health, prose, document, and quantitative literacy scales; a simple average percent correct would not allow reporting results that were comparable for all respondents. IRT models calculate the probability of answering a question correctly as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance on some latent trait can be compared across groups, such as those defined by sex, race/ethnicity, or place of birth.

IRT models assume that an examinee’s performance on each item reflects characteristics of the item and characteristics of the examinee. All models assume that all items on a scale measure a common latent ability or proficiency dimension (e.g., prose literacy) and that the probability of a correct response on an item is uncorrelated with the probability of a correct response on another item, given fixed values of the latent trait. Items are measured in terms of their difficulty as well as their ability to discriminate among examinees of varying ability.

The assessment used two types of IRT models to estimate scale scores. The two-parameter logistic (2PL) model was used for dichotomous items (that is, items that are scored either right or wrong). For the partial credit items, the graded response logistic (GRL) model was used. The scale indeterminacy was solved by setting an origin and unit size to the

reported scale means and standard deviations from the 1992 assessment. Linear transformation was performed to transform the original scale metric to the final reporting metric.

IRT models predict the probability of success on an item for each point along the latent ability scale. By selecting a criterion value for this probability, a single scale point can be associated with the difficulty of each item, and visual displays can be constructed showing the difficulty of selected items along the scale. Such item maps aid in interpreting the assessment scales and in describing the performance levels. The assessment conformed to common industry practice by choosing the value of 0.67 as its response probability convention.

5. DATA QUALITY AND COMPARABILITY

The NAAL sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest.

Sampling Error

In the 2003 survey, the use of a complex sample design, adjustments for nonresponse, and poststratification procedures resulted in dependence among the observations. Therefore, a jackknife replication method was used to estimate the sampling variance. The mean square error of replicate estimates around their corresponding full sample estimate provides an estimate of the sampling variance of the statistic of interest. The replication scheme was designed to produce stable estimates of standard errors for national and prison estimates as well as for the individual states.

The advantage of compositing the national and state samples during sample weighting was the increased sample size, which improved the precision of both the state and national estimates. However, biases could be present because the national PSU sample strata were not designed to maximize the efficiency of state-level estimates.

Nonsampling Error

The major source of nonsampling error in the 2003 NAAL was nonresponse error; special procedures were developed to minimize potential nonresponse bias based on how much of the survey the respondent completed. Other possible sources of nonsampling error were random measurement error and systematic error due to interviewers, coders, or scorers.

Coverage error. Coverage error could result from either the sampling frame of households or prisons being incomplete or from a household's or prison's failure to include all adults age 16 and older on the lists from which the sampled respondents were drawn. Special procedures and edits were built into NAAL to review both listers' and interviewers' ongoing work and to give any missed structures and/or dwelling units a chance of selection at data collection. However, just as all other household personal interview surveys have persistent undercoverage problems, the 2003 survey had problems in population coverage due to interviewers not gaining access to households in dangerous neighborhoods, locked residential apartment buildings, and gated communities.

Nonresponse error.

Unit nonresponse. Since three survey instruments—the screener, background questionnaire, and exercise booklet—were required for the administration of the survey, it was possible for a household or respondent to refuse to participate at the time of the administration of any one of these instruments. Because the screener and the background questionnaire were read to the survey participants in English or Spanish, but the exercise booklet required reading and writing in the English language, it was possible to complete the screener or background questionnaire but not the exercise booklet. Thus, response rates were calculated for each of the three instruments for the household samples. For the prison sample, there were only two points at which a respondent could not respond—at the administration of the background questionnaire or the exercise booklet.

For occupied households, "refusal or breakoff" was the most common explanation for nonresponse to the screener and the background questionnaire. The second most common explanation was "not at home after maximum number of calls." Nonresponse also resulted from language, physical, and mental problems. Housing units or individuals who refused to participate before any information was collected about them, or who did not answer a sufficient number of background questions, were not incorporated into the database. Because these individuals were unlikely to know that the survey intended to assess their literacy, it was assumed that their reason for not completing the survey was not related to their level of literacy.

There were reasons to believe that the literacy performance data were missing more often for adults with lower levels of literacy than for adults with higher levels. Field-test evidence and experience with surveys indicated that adults with lower levels of literacy were more likely than adults with higher levels either to decline to respond to the survey at all or to begin the

assessment but not complete it. Ignoring this pattern of missing data would have resulted in overestimating the literacy skills of adults in the United States. Therefore, to minimize bias in the proficiency estimates due to nonresponse to the literacy assessment, special procedures were developed to impute the literacy proficiencies of nonrespondents who completed fewer than five literacy tasks.

The household sample was subject to unit nonresponse from the screener, background questionnaire, literacy assessment, and oral module and to item nonresponse to background questionnaire items. Although all background questionnaire items had response rates of more than 85 percent, two stages of data collection—the screener and the background questionnaire—had unit response rates below 85 percent and thus required an analysis of the potential for nonresponse bias.

Table 13 presents a summary of the household response rate and table 14 presents a summary of the prison response rate.

Table 13. Weighted and unweighted unit response rates in the household sample of the National Assessment of Adult Literacy, by survey component: 2003

Component	Weighted response rate (percent)	Unweighted response rate (percent)
Screener	81.2	81.8
Background questionnaire	76.6	78.1
Literacy assessment	96.6	97.2
Overall response rate before imputation	60.1	62.1
Overall response rate after imputation	62.1	63.9

SOURCE: Greenberg, E., and Jin, Y. (2007). *2003 National Assessment of Adult Literacy: Public-Use Data File User's Guide* (NCES 2007-464). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Item nonresponse. For each background questionnaire, staff verified that certain questions providing critical information for weighting and data analyses had been answered, namely, education level, employment status, parents' level of education, race, and sex. If a response was missing, the case was returned to the field for data retrieval. Therefore, item response rates for completed background questionnaires were quite high, although they varied by type of question. Questions asking country of origin (first question in the booklet) and sex (last question in the booklet) had nearly 100 percent response rates, indicating that most respondents attempted to complete the entire questionnaire. Response rates were lower, however, for questions about income and educational background.

Table 14. Weighted and unweighted response rates in the prison sample of the National Assessment of Adult Literacy, by survey component: 2003

Component	Weighted response rate (percent)	Unweighted response rate (percent)
Prison	97.3	97.3
Background questionnaire	90.6	90.4
Literacy assessment	98.9	98.8
Overall response rate before imputation	87.2	86.8
Overall response rate after imputation	88.3	87.9

SOURCE: Greenberg, E., and Jin, Y. (2007). *2003 National Assessment of Adult Literacy: Public-Use Data File User's Guide* (NCES 2007-464). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

The CD-ROM: 2003 National Assessment of Adult Literacy Public-Use Data File User's Guide (Greenberg & Jin 2007) provides counts of item nonresponse. These, however, have to be considered in terms of the number of adults that were offered each task, because a great deal of the missing data is missing by design.

Nonresponse bias. NCES statistical standards require a nonresponse bias analysis when the unit response rate for a sample is less than 85 percent. The nonresponse bias analysis of the household sample revealed differences in the background characteristics of respondents who participated in the assessment compared with those who refused.

In bivariate unit-level analyses at the screener and background questionnaire stages, estimated percentages for respondents were compared with those for the total eligible sample to identify any potential bias owing to nonresponse. Although some statistically significant differences existed, the potential for bias was small because the absolute difference between estimated percentages was less than 2 percent for all domains considered. Multivariate analyses were conducted to further explore the potential for nonresponse bias by identifying the domains with the most differential response rates. These analyses revealed that the lowest response rates for the screener

were among dwelling units in segments with high median income, small average household size, and a large proportion of renters. The lowest response rates for the background questionnaire were among males age 30 and older in segments with high median income.

However, the variables used to define these areas and other pockets with low response rates were used in weighting adjustments. The analysis showed that weighting adjustments were highly effective in reducing the bias. The general conclusion was that the potential amount of nonresponse bias attributable to unit nonresponse at the screener and background questionnaire stages was likely to be negligible.

Measurement error. All background questions and literacy tasks underwent extensive review by subject area and measurement specialists, as well as scrutiny to eliminate any bias or lack of sensitivity to particular groups. Special care was taken to include materials and tasks that were relevant to adults of widely varying ages. During the test development stage, the tasks were submitted to test specialists for review, part of which involved checking the accuracy and completeness of the scoring guide. After preliminary versions of the assessment instruments were developed and after the field test was conducted, the literacy tasks were closely analyzed for bias or "differential item functioning." The goal was to identify any assessment tasks that were likely to underestimate the proficiencies of a particular subpopulation, whether it be older adults, females, or Black or Hispanic adults. Any assessment item that appeared to be biased against a subgroup was excluded from the final survey. The coding and scoring guides also underwent further revisions after the first responses were received from the main data collection.

Interviewer error checks. Several quality control procedures related to data collection were used during the field operation: an interviewer field edit, a complete edit of all documents by a trained field editor, validation of 10 percent of each interviewer's closeout work, and field observation of both supervisors and interviewers.

Coding/scoring error checks. In order to monitor the accuracy of coding, the questions dealing with country of birth, language, wages, and date of birth were checked in 10 percent of the questionnaires by a second coder. For the industry and occupation questions, 100 percent of the questionnaires were recoded by a second coder. Twenty percent of all the exercise booklets were subjected to a reader reliability check, which entailed a scoring by a second reader.

6. CONTACT INFORMATION

For content information about the NAAL project, contact:

Andrew Kolstad
Phone: (202) 502-7374
E-mail: andrew.kolstad@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street, NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Greenberg, E., and Jin, Y. (2007). *2003 National Assessment of Adult Literacy: Public-Use Data File User's Guide* (NCES 2007-464). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., and Dunleavy, E. (2007). *Literacy in Everyday Life: Results From the 2003 National Assessment of Adult Literacy* (NCES 2007-480). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kutner, M., Greenberg, E., Jin, Y., and Paulsen, C. (2006). *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy* (NCES 2006-483). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

White, S., and Dillow, S. (2006). *Key Concepts and Features of the 2003 National Assessment of Adult Literacy* (2006-471). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC

Chapter 21: Trends in International Mathematics and Science Study (TIMSS)

1. OVERVIEW

The Trends in International Mathematics and Science Study (TIMSS) is a study of classrooms across the country and around the world. The National Center for Education Statistics (NCES), in the Institute of Education Sciences at the U.S. Department of Education, is responsible for the implementation of TIMSS in the United States. Beginning in 1995 and every 4 years thereafter, TIMSS has provided participating countries with an opportunity to measure students' progress in mathematics and science achievement. Studies of students, teachers, schools, curriculum, instruction, and policy issues are also carried out to understand the educational context in which learning takes place.

TIMSS represents the continuation of a long series of studies conducted by the IEA. The IEA conducted its First International Mathematics Study (FIMS) in 1964 and the Second International Mathematics Study (SIMS) in 1980–82. The First and Second International Science Studies (FISS and SISS) were carried out in 1970–71 and 1983–84, respectively. Since the subjects of mathematics and science are related in many respects and since there is broad interest among countries in students' abilities in both subjects, TIMSS began to be conducted as an integrated assessment of both mathematics and science.

In 1995, TIMSS collected data on grades 3 and 4 as well as grades 7 and 8, and the final grade of secondary school (grade 12 in the United States), with 42 countries participating. In 1999, data were collected only for 8th-grade students, with 38 countries participating. For TIMSS 2003 and 2007, data were collected on grades 4 and 8, with 46 countries participating in 2003 and 58 countries participating in 2007.

In addition to the math and science assessments given to students, supplementary information is obtained through the use of student, teacher, and school questionnaires. Also, in 1995 and 1999, further component studies were implemented, including benchmark and video studies.

The TIMSS 1999 Benchmarking Study included states and districts or consortia of districts from across the United States that chose to participate. These states and districts completed the assessments and questionnaires following the same procedures developed for the participating countries. They then used the findings to assess their comparative international standing and to evaluate their mathematics and science programs in an international context.

For the TIMSS Videotape Study, designed as the first study to collect videotaped records of classroom instruction, representative samples of 8th-grade mathematics classes in 1995 and 1999 and science classes in 1999 were drawn and one lesson in each of the participating classrooms was videotaped. The analysis provides a more detailed context for understanding mathematics and science teaching and learning in the classroom.

WORLDWIDE STUDY OF CLASSROOMS WITH AS MANY AS 58 COUNTRIES PARTICIPATING

TIMSS tests a variety of subject and content areas:

- Grade 4 math: Number, geometric shapes and measures, data display
- Grade 8 math: Number, algebra, geometry, data and chance
- Grade 4 science: Life, physical, and Earth science
- Grade 8 science: Earth science, biology, chemistry, and physics

Purpose

TIMSS is designed to measure student performance in mathematics and science against what is expected to be taught in school. This focus on school curriculum allows for two broad questions to be addressed through TIMSS: (1) How do mathematics and science education environments differ across countries, how do student outcomes differ, and how are differences in these outcomes related to differences in mathematics and science education environments? (2) Are there patterns of relationships among contexts, inputs, and outcomes within countries that can lead to improvements in the theories and practices of mathematics and science education?

Components

TIMSS uses several types of instruments to collect data about students, teachers, schools, and national policies and practices that may contribute to student performance.

Written assessment. Assessments are developed to test students in various content areas within mathematics and science. For grade 4, the mathematics content areas are numbers; geometric shapes and measures; and data display. The grade 4 science content areas are Earth science; life science; and physical science. The grade 8 mathematics content areas are numbers; algebra; geometry; and data and chance. The grade 8 science content areas are biology; physics; chemistry; and Earth science.

In addition to being familiar with the mathematics and science content areas encountered in TIMSS, students are required to draw on a range of cognitive skills to successfully complete the assessment. TIMSS focuses on three cognitive domains in each subject: *knowing*, which covers the facts, procedures, and concepts students need to know; *applying*, which focuses on the ability of students to apply their knowledge and conceptual understanding to solve problems; and *reasoning*, which goes beyond solving routine problems to include unfamiliar situations and context that may require multi-step problem-solving.

After each TIMSS assessment cycle, approximately half of the items are publicly released, and replacement items that closely match the content of the original items are developed by international assessment and content experts. These new items are field tested and refined to the point where a variety of multiple choice and extended constructed-response items (i.e., items requiring written explanations from students) are chosen to be included in the TIMSS item pool.

Each student is asked to complete one booklet, made up of a subset of items taken from this item pool. No student answers all of the items in the item pool. The scoring of these booklets is accomplished through the use of a sophisticated and strict set of criteria that are implemented equally across all nations to ensure accuracy and comparability.

Student background questionnaire. Each student who takes the TIMSS assessment is asked to complete a questionnaire on issues including daily activities, family attributes, educational resources in the home, engagement in and beliefs about learning, instructional processes in the classroom, study habits, and homework.

Teacher questionnaire. The teacher questionnaire is given to the mathematics and science teachers of the students assessed in the study. These questionnaires ask about topics such as attitudes and beliefs about teaching and learning, teaching assignments, class size and organization, topics covered in class, the use of various teaching tools, instructional practices, professional preparation, and continuing development.

The teacher questionnaire is designed to provide information about the teachers of the students in the TIMSS student samples. The teachers who complete TIMSS questionnaires do not constitute a sample from any definable population of teachers. Rather, they represent the teachers of a national sample of students.

School questionnaire. The principal or head administrator is also asked to complete a questionnaire for the school focused on community attributes, personnel, teaching assignments, policy and budget responsibilities, curriculum, enrollment, student behavior issues, instructional organization, and mathematics and science courses offered.

Information collected from students, their teachers and schools is summarized in composite indices focused, in particular, on the relationship between mathematics and science achievement and the home, classroom, and school environment.

Curriculum questionnaire. The national research coordinator, or representative, of each participating country is asked to complete a questionnaire focused on the policies and practices supported at the national level that may contribute to student performance. In addition, because the mathematics and science topics covered in the assessment may not be included in all countries' curriculum, the national research coordinators are asked to indicate whether each topic

covered in TIMSS is included in their countries' intended curriculum through the fourth or eighth grade.

Encyclopedia. Beginning with TIMSS 2007, each participating country is asked to provide a written overview of the context in which mathematics and science instruction takes place, summarizing the structure of the education system, the mathematics and science curricula and instruction in primary and secondary grades, teacher education requirements, and the types of examinations and assessments employed to monitor success. The resulting chapters are compiled in *a publication entitled the TIMSS Encyclopedia*.

Videotape study. The 1995 TIMSS Videotape Study was designed as the first study to collect videotaped records of classroom instruction from national probability samples in Japan, Germany, and the United States in order to gather more in-depth information about the context in which learning takes place as well as to enhance understanding of the statistical indicators available from the main TIMSS study. An hour of regular classroom instruction was videotaped in a subsample of 8^{th}-grade mathematics classrooms (except in Japan, where videotaping was usually done in a different class, selected by the principal) included in the assessment phase of TIMSS in each of the three countries.

The 1999 TIMSS Videotape Study was expanded in scope to examine national samples of 8^{th}-grade mathematics and science instructional practices in seven nations: Australia, the Czech Republic, Hong Kong, Japan, the Netherlands, Switzerland, and the United States. Four countries—Australia, the Czech Republic, the Netherlands, and the United States—participated in both the mathematics and science components of the study. Hong Kong and Switzerland participated in only the mathematics component, and Japan in only the science component.

Curriculum studies. Continuing the approach of previous IEA studies, TIMSS addressed three conceptual levels of curriculum in 1995. The intended curriculum was composed of the mathematics and science instructional and learning goals as defined at the system level. The implemented curriculum was the mathematics and science curriculum as interpreted by teachers and made available to teachers. The attained curriculum was the mathematics and science content that students had learned and their attitudes toward these subjects. To aid in interpretation and comparison of results, TIMSS also collected extensive information about the social and cultural contexts for learning, many of which are related to variations among the education systems.

To gather information about the intended curriculum, mathematics and science specialists within each participating country worked section by section through curriculum guides, textbooks, and other curricular materials to categorize aspects of these materials in accordance with detailed specifications derived from TIMSS mathematics and science curriculum frameworks.

To collect data about how the curriculum was implemented in classrooms, TIMSS administered a broad array of questionnaires. These questionnaires were administered at the country level on decision making and organizational features within the education systems. The students who were tested answered questions pertaining to their attitudes toward mathematics and science, classroom activities, home background, and out-of-school activities. The mathematics and sciences teachers of sampled students responded to questions about teaching emphasis on the topics in the curriculum frameworks, instructional practices, textbook use, professional training and education, and their views on mathematics and science. The heads of schools responded to questions about school staffing and resources, mathematics and science course offerings, and support for teachers.

Ethnographic case studies. The case studies approach to understanding cultural differences in behavior has a long history in selected social science fields. Conducted only in 1995, the case studies were designed to focus on four key topics that challenge U.S. policymakers and to investigate how these topics were dealt with in the United States, Japan, and Germany: implementation of national standards; the working environment and training of teachers; methods for dealing with differences in ability; and the role of school in adolescents' lives. Each topic was studied through interviews with a broad spectrum of students, parents, teachers, and educational specialists. The ethnographic approach permitted researchers to explore the topics in a naturalistic manner and to pursue them in greater or lesser detail, depending on the course of the discussion. As such, these studies both validated and integrated the information gained from official sources with that obtained from teachers, students, and parents in order to ascertain the degree to which official policy reflected actual practice. The objective was to describe policies and practices in the nations under study that were similar to, different from, or nonexistent in the United States.

In three regions in each of the three countries, the research plan called for each of the four topics to be studied in the 4^{th}, 8^{th}, and 12^{th} grades. The specific cities and schools were selected "purposively" to represent

different geographical regions, policy environments, and ethnic and socioeconomic backgrounds. Schools in the case studies were separated from schools in the main TIMSS sample. Where possible, a shortened form of the TIMSS test was administered to the students in the selected schools. The ethnographic researchers in each of the countries conducted interviews and obtained information through observations in schools and homes. Both native-born and nonnative researchers participated in the study to ensure a range of perspectives.

TIMSS benchmarking study. In 1999, 13 states and 14 districts or consortia of districts throughout the United States participated as their own "nations" in this project, following the same guidelines as the participating countries. The samples drawn for each of these states and districts were representative of the student population in each of these states and districts. The findings from this project allowed these jurisdictions to assess their comparative international standing and judge their mathematics and science programs in an international context.

NAEP/TIMSS linking study. A subsample of students who took the 2000 state National Assessment of Educational Progress (NAEP) mathematics and science assessment also took the 1999 TIMSS assessment. (See chapter 18 for more information on NAEP.) This provided an opportunity to compare students' performance on NAEP to their performance on TIMSS, and allowed for estimates of how states participating in the 2000 NAEP would have performed had they participated in TIMSS 1999. Results from the TIMSS 1999 Benchmarking Study were used to check the results of the linking study.

Periodicity

First conducted in 1995, TIMSS has been conducted every 4 years since then. Previous international math studies were conducted in 1964 and 1980–82; previous international science studies were conducted in 1970–71 and 1983–84.

2. USES OF DATA

The possibilities for specific research questions to be dealt with by TIMSS are numerous; however, the main research questions, focusing on the student, the school or classroom, and the national or international levels, are illustrated below:

- How much mathematics and science have students learned?
- How well are students able to apply mathematics and science knowledge to problem solving?
- What are students' attitudes toward mathematics and science?
- What do teachers teach in their classrooms?
- What methods and materials do teachers use in teaching mathematics and science, and how are they related to student outcomes?
- How strongly are students motivated to learn, in general, and to the learning of mathematics and science, in particular?
- What factors characterize the academic and professional preparation of teachers of mathematics and science?
- What are teachers' beliefs and opinions about the nature of mathematics and science (and about teaching them), and how are they related to the comparable opinions and attitudes of their students?
- What methods do teachers use to evaluate their students?
- If there are national curricula in a country, how specific are they, and what efforts are made to see that they are followed?

3. KEY CONCEPTS

Key terms related to TIMSS are described below.

National Desired Population. The stated objective in TIMSS is that the National Desired Population within each country be as close as possible to the International Desired Population, which is the target population. (See "Target Population" below under Section 4. Survey Design.) Using the International Desired Population as a basis, participating countries have to operationally define their populations for sampling purposes. Some national research coordinators have to restrict coverage at the country level, for example, by excluding remote regions or a segment of their country's education system. Thus, the National Desired Population sometimes differs from the International Desired Population.

4. SURVEY DESIGN

National Research Coordinators. This is an official from each participating country appointed to implement national data collection and processing in accordance with international standards. In addition to selecting the sample of students, national research coordinators are responsible for working with school coordinators, translating the test instruments, assembling and printing the test booklets, and packing and shipping the necessary materials to the sampled schools. They are also responsible for arranging the return of the testing materials from the school to the national center, preparing for and implementing the constructed-response item scoring, entering the results into data files, conducting on-site quality assurance observations for a 10 percent sample of schools, and preparing a report on survey activities.

Target Population

The International Desired Population for all countries is defined as follows:

- Grade 4: All students enrolled in the grade that represents 4 years of schooling, counting from the 1st year of the International Standard Classification of Education (ISCED) Level 1, providing that the mean age at the time of testing is at least 9.5 years. For most countries, the target grade should be the fourth grade or its national equivalent. All students enrolled in the target grade, regardless of their age, belong to the international desired target population.

- Grade 8: All students enrolled in the grade that represents 8 years of schooling, counting from the 1st year of ISCED Level 1, providing that the mean age at the time of testing is at least 13.5 years. For most countries, the target grade should be the eighth grade or its national equivalent. All students enrolled in the target grade, regardless of their age, belong to the international desired target population.

Thus, TIMSS uses a grade-based definition of the target population.

Sample Design

Each country participating in TIMSS, like the United States, is required to draw random samples of schools. In the United States, a national probability sample is drawn for each study that has resulted in over 500 schools and approximately 33,000 students participating in 1995, approximately 220 schools and 9,000 students participating in 1999, approximately 480 schools and almost 19,000 students in 2003, and approximately 500 schools and over 20,000 students in 2007. This sample design ensures the appropriate number of schools and students are participating to provide a representative sample of the students in a specific grade in the United States as a whole.

The TIMSS sample design for each country and population is intended to give a probability sample of all students within the target grades in the national school system (except for a small number of students allowed to be excluded as ineligible according to national criteria). Every eligible student in the country's school system has a chance of being selected, with a fixed probability of selection. These probabilities of selection are designed to be equal across eligible students as much as possible, but for a variety of reasons the probabilities of selection differ between students in most of the national samples.

Written assessment.

The TIMSS sample design is a two-stage stratified cluster sample, with schools as the first stage of selection and classrooms within schools as the second stage of selection. For the first time TIMSS 2007 included an optional third stage. The third-stage sampling units for TIMSS 2007 were students within sampled classrooms. Generally however, TIMSS chooses intact classrooms, so students are essentially chosen at the same stage as the classroom (i.e. the second stage).

Individual schools are selected with probability proportionate to size (PPS), size being the estimated number of students enrolled in the target grade. Prior to sampling, schools in the sampling frame can be assigned to a predetermined number of explicit or implicit strata. Substitution schools, selected to replace schools that refuse to participate, are identified simultaneously.

The classroom sampling design is intended to be an equal probability design with no subsampling in the classroom. However, a design based on a PPS sample of classrooms, with a fixed sample size of students selected within the sampled classroom, is permitted under the international guidelines. Exclusions can occur at the school level, the classroom level, or the student level. TIMSS participants are expected to keep such exclusions to no more than 10 percent of the National Desired Population.

The optional third-stage sampling unit for TIMSS 2007 was students within the sampled classrooms. While all students in a sampled classroom were to be selected for the assessment, it was possible for participating

countries to sample a subgroup of students after consultation with Statistics Canada, the organization serving as the sampling referee.

TIMSS standards for sampling precision require a minimum of 4,000 students to be assessed per grade. To meet the standard, at least 150 schools are selected per target population. However, the clustering effect of sampling classrooms rather than students is also considered in determining the overall sample size. Because the magnitude of the clustering effect is determined by the size of the cluster and the intraclass correlation, TIMSS produced sample-design tables showing the number of schools to sample for a range of intraclass correlations and minimum-cluster-size values. Some countries need to sample more than 150 schools. Countries, however, are asked to sample 150 schools even if the estimated number of schools necessary to be sampled is less than 150.

The schools in each explicit stratum (geographical region, public/private, etc.) are listed in order of the implicit stratification variables and then further sorted according to their measure of size. The stratification variables differ from country to country. Small schools are handled either through explicit stratification or through the use of pseudo-schools. In some very large countries, there is a preliminary sampling stage before schools are sampled in which the country is divided into primary sampling units.

In cases where a sampled school is unable to participate in the assessment, a replacement school is used. The replacement school is the next school on the ordered school-sampling list as the replacement for each particular sampled school. The school after that is a second replacement, should it be necessary. Using either explicit or implicit stratification variables and ordering of the school sampling frame by size ensures that any original sampled school's replacement has similar characteristics.

In the second stage of sampling, classrooms of students are sampled. Generally, in each school, one classroom is sampled from each target grade, although some countries opt to sample two classrooms at the upper grade in order to be able to conduct special analyses. Most countries test all students in selected classrooms, and in these instances the classrooms are selected with equal probabilities. A few participants use a design based on a PPS sample of classrooms, with a fixed sample size of students selected within the sampled classrooms. Participants with particularly large classrooms in their schools can decide to subsample a fixed number of students from each selected classroom. This is done using a simple random sampling method whereby all students in a sampled classroom are assigned equal selection probabilities.

In the United States, TIMSS 2007 used a two-stage stratified cluster sampling design based on the 2006 NAEP school sampling frame. The United States did not use the optional third stage of sampling (i.e. students within classrooms) for TIMSS 2007. (Time constraints related to recruitment activities required sample selection before the 2007 frame became available.) For this purpose the sampling frame, though not explicitly stratified, was implicitly stratified by four categorical variables: type of school (public or private); region of the country (Northeast, Central, West, Southeast); community type (eight levels); and percentage of Black, Hispanic, and other race/ethnicity students (above or below 15 percent of the student population).

The first stage of the design used a systematic PPS technique to select schools for the original sample. That is, schools were selected with a probability proportionate to the school's estimated enrollment of fourth- or eighth-grade students. Enrollment data for public schools were taken from the 2003–04 Common Core of Data (CCD), and data for private schools were taken from the 2003–04 Private School Universe Survey (PSS). For each original school selected, the two adjacent schools in the sampling frame, and within the same implicit stratum, were designated as the first and second replacement schools. The first substitute followed the original sample school in the frame listing and the second substitute preceded it. Substitute schools were designed to be used only if an original school refused to participate. In this situation the first substitute was to be contacted first, with the second substitute contacted only if the first substitute also refused to participate. Additionally, one sampled school was not allowed to substitute for another, and a given school could not be assigned to substitute for more than one sampled school.

An initial sample of 300 schools was selected at each grade level. Ineligible schools among these reduced the grade 4 sample to 290 schools and the grade 8 sample to 290 schools.

At each grade level, the U.S. sample design within schools consisted of an equal probability sample of two classrooms. In schools with a single eligible classroom, that classroom was selected with certainty. All eligible students in the classroom were designated to be in the sample (although generally the option for sub sampling did exist, there was no subsampling of students in the TIMSS 2007 U.S. sample).

Teacher questionnaire. The TIMSS database for each country includes questionnaire data from the teachers of the sampled classrooms, which can be linked to student assessment data in the classrooms. Any teacher linked as mathematics or science teacher to any assessed student is eligible to receive a questionnaire. The classroom sample is drawn from a listing of mathematics classrooms, so that in most situations only one mathematics teacher is linked to each sampled classroom. If this single teacher is also only linked to a single sampled classroom, then the teacher receives a questionnaire for that single classroom.

This straightforward one-to-one linking does not always hold, however. In some cases, teachers may teach both mathematics and science to students in a sampled classroom, making them eligible to receive questionnaires for both subjects.

For the U.S. TIMSS 2007 sample, a teacher was not asked to complete more than one questionnaire. In cases where a teacher taught both subject areas, the teacher was provided a specially designed questionnaire that included questions for both mathematics and science teachers.

In general, each country is allowed to develop its own methodology for this process of assigning subjects and classrooms to teachers when the links are not straightforward due to the presence of one to many (or many to one) mappings.

Assessment Design

TIMSS is a cooperative effort involving representatives from every country participating in the study. For TIMSS 2007, the development effort began with a revision of the frameworks that were used to guide the construction of the assessment. The frameworks were updated to reflect changes in the curriculum and instruction of participating countries. Extensive input from experts in mathematics and science education, assessment, and curriculum, and representatives from national education centers around the world contributed to the final shape of the frameworks used in 2007. Maintaining the ability to measure change over time is an important factor in constantly revising the frameworks.

Test development. As part of the TIMSS dissemination strategy, approximately one-half of the items at each grade are released for public use. To replace assessment items that have been released, countries submit items for review by subject-matter specialists, and additional items are written to ensure that the content, as explicated in the frameworks, is covered adequately. Items are reviewed by an international Science and Mathematics Item Review Committee and field tested in most of the participating countries. Results from the field tests are used to evaluate item difficulty, how well items discriminate between high- and low-performing students, the effectiveness of distracters in multiple-choice items, scoring suitability and reliability for constructed-response items, and evidence of bias toward or against individual countries or in favor of boys or girls.

Instrument design. TIMSS 2007 included booklets containing assessment items as well as questionnaires submitted to principals, teachers, and students. The assessment booklets were constructed such that not all of the students responded to all of the items, which is consistent with the design of other large-scale assessments, such as NAEP. To keep the testing burden to a minimum, and to ensure broad subject-matter coverage, TIMSS 2007 used a rotated block design that included both mathematics and science items. That is, students encountered both mathematics and science items during the assessment.

The U.S. 2007 fourth-grade assessment consisted of 14 booklets, each requiring approximately 72 minutes of response time. The 14 booklets were rotated among students, with each participating student completing only 1 booklet. The mathematics and science items were assembled into 14 blocks, or clusters, of items, with each block containing either mathematics or science items. The secure, or trend, items were included in 3 blocks, with the other 11 blocks containing replacement items. Each of the 14 booklets contained a total of 6 blocks.

The U.S. 2007 eighth-grade assessment consisted of 18 booklets, each requiring approximately 90 minutes of response time. The 18 booklets were rotated among students, with each participating student completing only 1 booklet. The mathematics and science items were assembled into 14 blocks, or clusters, of items, with each block containing either mathematics or science items. The secure, or trend, items were included in 3 blocks, with the other 11 blocks containing replacement items. Each of the 18 booklets contained a total of 4 blocks. As part of the design process, it was necessary to ensure that the booklets showed a distribution across the mathematics and science content domains as specified in the frameworks.

Data Collection and Processing

Data collection. TIMSS 2007 emphasized the use of standardized procedures in all countries. Each country collected its own data, based on comprehensive manuals and trainings provided by the international

project team to explain the survey's implementation, including precise instructions for the work of school coordinators and scripts for test administrators to use in testing sessions. Test administration in the United States was carried out by professional staff trained according to the international guidelines. School staff was asked only to assist with listings of students, identifying space for testing in the school, and specifying any parental consent procedures needed for sampled students.

Each country was responsible for conducting quality control procedures and describing this effort in the national research coordinator's report documenting procedures used in the study. In addition, the TIMSS International Study Center considered it essential to monitor compliance with the standardized procedures. National research coordinators were asked to nominate one or more persons unconnected with their national center, such as retired school teachers, to serve as quality control monitors for their countries. The International Study Center developed manuals for the monitors and briefed them in 2-day training sessions about TIMSS 2007, the responsibilities of the national centers in conducting the study, and their own roles and responsibilities.

Data entry and cleaning. Responsibility for data entry is taken by the national research coordinator from each participating country. The data collected for TIMSS 2007 were entered into data files with a common international format, as specified in the *Manual for Entering the TIMSS 2007 Data*. Data entry was facilitated by the use of common software available to all participating countries (WinDEM). The software facilitated the checking and correction of data by providing various data consistency checks. After data entry, the data were sent to the IEA Data Processing Center (DPC) in Hamburg, Germany, for cleaning. The DPC checked that the international data structure was followed; checked the identification system within and between files; corrected single-case problems manually; and applied standard cleaning procedures to questionnaire files. Results of the data cleaning process were documented by the DPC. This documentation was then shared with the national research coordinator with specific questions to be addressed. The national research coordinator then provided the DPC with revisions to coding or solutions for anomalies. The DPC then compiled background univariate statistics and preliminary classical and Rasch Item Analysis.

Estimation Methods

Once TIMSS data are scored and compiled, the responses are weighted according to the sample design and population structure and then adjusted for nonresponse. This ensures that countries' representation in TIMSS is accurately assessed. The analyses of TIMSS data for most subjects are conducted in two phases: scaling and estimation. During the scaling phase, Item Response Theory (IRT) procedures are used to estimate the measurement characteristics of each assessment question. During the estimation phase, the results of the scaling are used to produce estimates of student achievement (proficiency) in the various subject areas. The methodology of multiple imputations (plausible values) is then used to estimate characteristics of the proficiency distributions. Although imputation is conducted for the purpose of determining plausible values, no imputations are included in the TIMSS database.

Weighting. The TIMSS international design provides for two categories of sampling weights. The first category is designed to be used when schools, classrooms, or students are the unit of analysis. The second category is designed to be used in analyses where teachers, or both teachers and students, are the units of analysis.

First category. Sampling weights in the first category consist of school, classroom, and student weights, along with a combined student weight that is the product of these weights. The school weight is, essentially, the inverse of the probability of a school being sampled in the first stage of the sampling design. A school-level nonresponse adjustment is applied to compensate for any sampled schools that did not participate and were not replaced. This adjustment is calculated independently for each explicit stratum.

Classroom weights reflect the probability of the sampled classroom(s) being selected from among all the classrooms in the school at the target grade level. This classroom weight is calculated independently for each participating school. If a sampled classroom in a school does not participate, or if the participation rate among students in a classroom falls below 50 percent, a classroom-level participation adjustment is made to the classroom weight. If one (or more) selected classrooms in a school do not participate, the classroom participation adjustment is computed at the explicit stratum level rather than at the school level to reduce the risk of bias.

In the first category, student sampling weights are set at 1.0 since intact classrooms are sampled and each student in the sampled classrooms is certain of selection. A nonresponse adjustment is applied to adjust for sampled students who do not take part in the testing. This adjustment is calculated independently for each sampled classroom. An overall student sampling

weight is provided as well and is calculated as the product of the school, class, and student weights described above.

In addition, TIMSS provides “house” and “senate” weights, which are scaled versions of the overall student weight just described. The names are derived from an analogy with the U.S. legislative system. House weights are a set of weights based on the total sample size of each country, to be used when estimates across countries are computed or significance tests performed. The transformation of the weights will be different within each country, but in the end, the sum of the house-weight variables within each country will total to the sample size for that country. The house-weight variable is proportional to the total weight for that variable by the ratio of the sample size divided by the size of the population. These sampling weights can be used when the data user wants the actual sample size to be used in performing significance tests.

Senate weights are a set of weights based on a constant scalar, to be used when estimates across countries are computed or significance tests performed. The transformation of the weights will be different within each country, but in the end, the sum of the senate-weight variables within each country will total to a fixed value. The senate-weight variable, within each country, is proportional to the total weight for that variable by the ratio of the fixed value divided by the size of the population estimate. These sampling weights can be used when cross-national comparisons are required and the data user wants to have each country contribute the same amount to the comparison, regardless of the size of the population.

Second category. The teacher weight is a teacher-classroom weight and so is greater than 0 for a classroom only if the teacher filled out a questionnaire for that classroom. The teacher-classroom weight is equal to the sum of the student-teacher weights (see discussion below) for students linked to a classroom for a particular assessment.

Sampling weights in this second category are provided to facilitate analyses in which student and teacher data are analyzed together. TIMSS does not provide for a sample of teachers. Rather, the teachers in question are those who teach the sample of TIMSS students. As a consequence, analyses involving teachers have to be viewed as student-level analyses. Accordingly, teacher weights and student-teacher weights are derived from the overall student weight and are designed to accommodate the fact that students may have more than one teacher. Teacher weights are calculated by dividing the sampling weight for a student by the number of teachers that the student has. Separate mathematics and science student-teacher weights are developed by dividing the student sampling weight by, respectively, the number of mathematics teachers and the number of science teachers that the student has.

Scaling. TIMSS 1995, 1999, 2003, and 2007 used IRT procedures to produce scale scores that summarized the achievement results. With this method, the performance of a sample of students in a subject area or subarea can be summarized on a single scale or a series of scales, even when different students are administered different items. Because of the reporting requirements for TIMSS and because of the large number of background variables associated with the assessment, a large number of analyses have to be conducted. The procedures TIMSS uses for the analyses are developed to produce accurate results for groups of students while limiting the testing burden on individual students. Furthermore, these procedures provide data that can be readily used in secondary analyses. IRT scaling provides estimates of item parameters (e.g., difficulty, discrimination) that define the relationship between the item and the underlying variable measured by the test. IRT model parameters are estimated for each test question, with an overall scale being established as well as scales for each predefined content area specified in the assessment framework. For example, the TIMSS 2007 8th-grade mathematics assessment had four scales describing mathematics content strands, and the science assessment had scales for four fields of science.

Imputation and plausible values. Although multiple imputation techniques are applied to create plausible values for student proficiency scores, with one exception, imputations were not generated for missing values in the TIMSS 2007 teacher, school, or student questionnaire data files. The single exception refers to a U.S.-only variable in the school file, the principal’s report of the percentage of students eligible for free- or reduced-price lunch. For public schools, missing values for this variable were replaced by information obtained from the CCD. Analogous information was not available for private schools. Subsequently, analyses were undertaken to ensure that confidentiality was maintained.

During the scaling phase, plausible values are used to characterize scale scores for students participating in the assessment. To keep student burden to a minimum, TIMSS administers a limited number of assessment items to each student; too few to produce accurate content-related scale scores for each student. To account for this, for each student, TIMSS generates five possible content-related scale scores that represent

selections from the distribution of content-related scale scores of students with similar backgrounds who answer the assessment items the same way. The plausible-values technology is one way to ensure that the estimates of the average performance of student populations and the estimates of variability in these estimates are more accurate than those determined through traditional procedures, which estimate a single score for each student.

While constructing plausible values, careful quality control steps ensure that the subpopulation estimates based on these plausible values are accurate. Plausible values are constructed separately for each national sample. TIMSS uses the plausible-values methodology to represent what the true performance of an individual might have been, had it been observed. This is done by using a small number of random draws from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. The draws from the distribution are different from one another to quantify the degree of precision (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performance. The TIMSS plausible values function like point estimates of scale scores for many purposes, but they are unlike true point estimates in several respects. They differ from one another for any particular student, and the amount of difference quantifies the spread in the underlying distribution of possible scale scores for that student. Because of the plausible-values approach, secondary researchers can use the TIMSS data to carry out a wide range of analyses.

Scale anchoring. Beginning with TIMSS 2003, the percentage of students in each country performing at each of four international benchmarks of performance are reported. The benchmarks are selected to represent the range of performance of students internationally. The four benchmarks selected to represent points along the scale are advanced (set at 625), high (550), intermediate (475), and low (400). Using these points along the TIMSS scale, a scale anchoring analysis is conducted to describe student performance in terms of what they know and can do. The scale anchoring process involves a statistical component, which identifies assessment items that discriminate between points on the scale, and expert judgment, in which subject-matter specialists examine the items that anchor at different points along the scale and generalize about students' knowledge and understanding.

Future Plans

The next TIMSS data collection will take place in spring 2011. In addition, a new effort to link national and international assessments will be initiated in 2011 so that states can compare their own students' performance against international benchmarks. The linking study is intended to enable NCES to project state-level scores on the TIMSS using data from the National Assessment of Educational Progress (NAEP).

In the linking study, two representative national samples will be tested on their knowledge of mathematics and science by taking both the NAEP and TIMSS assessments. One sample of 10,000 eighth-graders will take combined test booklets in the winter of 2011 as part of NAEP. The other sample of 7,500 eighth-graders will take combined test booklets in the spring of 2011 as part of TIMSS. The relationships between the two assessments of mathematics and science that are found in these two samples will permit state-level projections of how the students in the 50 states and the District of Columbia that took NAEP would have performed in eighth-grade mathematics and science on TIMSS, with scores that can be compared to those of other countries. Data from a number of states that have agreed to administer TIMSS 2011 to state representative samples will be compared to the projected scores to ensure the accuracy of the linking projections.

5. DATA QUALITY AND COMPARABILITY

In addition to setting high standards for data quality, the TIMSS International Study Center has tried to ensure the overall quality of the study through a dual strategy of providing support to the national centers and performing quality control checks.

Despite the efforts taken to minimize error, any sample survey as complex as TIMSS has the possibility of error. Below is a discussion of possible sources of error in TIMSS.

Sampling Error

With complex sampling designs that involve more than the simple random sampling of students, as in the case of the stratified multistage design used in TIMSS 2007, where students were clustered within schools, there are several methods for estimating the sampling error of a

statistic that avoid the assumption of simple random sampling. One such method is the Jackknife Repeated Replication (JRR) technique. The particular application of the JRR technique used in TIMSS is termed a paired selection model because it assumes that the primary sampling units can be paired in a manner consistent with the sampling design, with each pair regarded as members of a pseudo-stratum for variance estimation purposes.

Following this first-stage sampling, there may be any number of subsequent stages of selection that may involve equal or unequal probability selection of the corresponding elements.

Imputation error. The variance introduced by imputation of missing data must be considered when using plausible values to estimate standard errors for proficiency estimates. The general procedure for estimating the imputation variance using plausible values is as follows: first estimate the statistic (t), each time using a different set of the plausible values (M). The statistics t_m can be anything estimable from the data, such as a mean, the difference between means, percentiles, etc. If all five plausible values in the TIMSS database are used, the parameter will be estimated five times, once using each set of plausible values. Each of these estimates will be called t, where $m=1, 2,\ldots, 5$. Once the statistics are computed, the imputation variance is then computed as

$$Var_{imp} = (1 + 1/M)Var(t_m)$$

where M is the number of plausible values used in the calculation, and $Var(t_m)$ is the variance of the estimates computed using each plausible value.

Nonsampling Error

Due to the particular situations of individual TIMSS countries, sampling and coverage practices have to be adaptable, in order to ensure an internationally comparable population. As a result, nonsampling errors in TIMSS can be related both to coverage error and nonresponse. Measurement error is also a nontrivial issue in administering TIMSS, as different countries have different mathematics and science curricula. These potential sources of error are discussed in detail below.

Coverage error. The stated objective in TIMSS is that the effective population, the population actually sampled by TIMSS, be as close as possible to the International Desired Population. Yet, because a purpose of TIMSS is to study the effects of different international curricula and pedagogical methods on mathematics and science learning, participating countries have to operationally define their population for sampling purposes. Some national research coordinators have to restrict coverage at the country level, for example, by excluding remote regions or a segment of their country's education system. In these few situations, countries are permitted to define a National Desired Population that does not include part of the International Desired Population. Exclusions can be based on geographic areas or language groups.

Nonresponse error. Unit nonresponse error results from nonparticipation of schools and students. Weighted and unweighted response rates are computed for each participating country by grade, at the school level, and at the student level. Overall response rates (combined school and student response rates) are also computed.

The minimum acceptable school-level response rate for all countries, before the use of replacement schools, is set at 85 percent. This criterion is applied to the unweighted school-level response rate. However, both weighted and unweighted school-level response rates are calculated, with and without replacement schools. It is generally the case that weighted and unweighted response rates are similar.

Like the school-level response rate, the minimum acceptable student-level response rate is set at 85 percent for all countries. This criterion is applied to the unweighted student-level response rate. However, both weighted and unweighted student-level response rates are calculated. The weighted student-level response rate is the sum of the inverse of the selection probabilities for all participating students divided by the sum of the inverse of the selection probabilities for all eligible students.

Table 15 shows the unweighted unit level response rates for the data collections of 1995, 1999, 2003, and 2007 for grades 4 and 8.

Measurement error. Measurement error is introduced into a survey when its test instruments do not accurately measure the knowledge or aptitude they are intended to assess. The largest potential source of measurement error in TIMSS results from differences in the mathematics and science curricula across participating countries. In order to minimize the effects of measurement error, TIMSS carries out a special test called the *Test-Curriculum Matching Analysis*. Each country is asked to identify, for each item, whether the topic of the item is in the curriculum of the majority of the students.

Data Comparability

Through a careful process of review, analysis, and refinement, the assessment and questionnaire items are purposefully developed and field tested for similarity and for reliable comparisons between survey years. After careful review of all available data, including a test for reliability between old and new items, the TIMSS assessments are found to be very similar in format, content, and difficulty level across years.

Table 15. TIMSS unweighted unit-level response rates, by level, year, and grade: 1995, 1999, 2003, and 2007

Year and grade	School	Student	Overall
1995			
4th grade	86	94	81
8th grade	84	92	77
1999			
4th grade	†	†	†
8th grade	90	93	84
2003			
4th grade	83	95	78
8th grade	78	94	73
2007			
4th grade	89	95	84
8th grade	83	93	77

† Not available. TIMSS did not collect data from grade 4 in 1999.

SOURCE: Martin, M.O., and Kelly, D.L. (Eds.). (1998). *TIMSS Technical Report: Volume II: Implementation and Analysis, Primary and Middle School Years*. Boston College, International Study Center. Chestnut Hill, MA. Martin, M.O., Gregory, G.D., and Stemler, S.E. (Eds.). (2000). *TIMSS 1999 Technical Report*. Boston College, International Study Center. Chestnut Hill, MA. Martin, M.O., Mullis, I.V.S., and Chrostowski, S.J. (Eds.). (2004). *TIMSS 2003 Technical Report: Findings From IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades*. Boston College, International Study Center. Chestnut Hill, MA. Olson, J.F., Martin, M.O., and Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 Technical Report*. Boston College, International Study Center. Chestnut Hill, MA.

Findings from comparisons between the results of TIMSS, however, cannot be interpreted to indicate the success or failure of mathematics and science reform efforts within a particular country, such as the United States. International experts develop the TIMSS curriculum frameworks to portray the structure of the intended school mathematics and science curricula from many nations, not specifically the United States. Thus, when interpreting the findings, it is important to take into account the mathematics and science curricula likely encountered by U.S. students in school. TIMSS results are most useful, however, when they are considered in light of knowledge about education systems that include curricula, but also factors in trends in education reform, changes in school-age populations, and societal demands and expectations.

The ability to compare data across different countries constitutes a considerable part of the purpose behind TIMSS. As a result, it is crucial to ensure that items developed for use in one country are functionally identical to those used in other countries. Because questionnaires are originally developed in English and later translated into the language of each of the TIMSS countries, some differences do exist in the wording of questions. National research coordinators from each country review the national adaptations of individual questionnaire items and submit a report to the IEA Data Processing Center. In addition to the translation verification steps used for all TIMSS test items, a thorough item review process is used to further evaluate any items that are functioning differently in different countries according to the international item statistics. In certain cases, items have to be recoded or deleted entirely from the international database as a result of this review process.

6. CONTACT INFORMATION

For content information about the TIMSS project, contact:

Patrick Gonzales
Phone: (415) 920-9229
E-mail: patrick.gonzales@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

Most of the technical documentation for TIMSS is published by the International Study Center at Boston College. The U.S. Department of Education, National Center for Education Statistics, is the source of several additional references listed below; these publications are indicated by an NCES number.

General

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W. (2000). *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999* (NCES 2001-028). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Gonzales, P., Guzmán, J.C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., and Williams, T. (2004). *Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005-005). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Uses of Data

Johnson, E.G. (1998). *Linking the National Assessment of Educational Progress (NAEP) and The Third International Mathematics and Science Study (TIMSS): A Technical Report (*NCES 98-499). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Johnson, E.G., and Siegendorf, A. (1998). *Linking the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): Eighth-Grade Results (*NCES 98-500). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kastberg, D., Roey, S., Jocelyn, L., and Williams, T. (2006). *Trends in International Mathematics and Science Study (TIMSS) 2003 U.S. Datafile and User's Guide (*NCES 2006-058). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., Smith, C., and Stearns, P. (2009). *TIMSS 2007 U.S. Technical Report and User Guide* (NCES 2009-012). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Rutkowski, L., González, E., Joncas, M., and von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher 39* (2), 142-151.

Survey Design

Martin, M.O., and Kelly, D.L. (Eds.). (1996). *TIMSS Technical Report: Volume I: Design and Development*. Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., and Kelly, D.L. (Eds.). (1998). *TIMSS Technical Report: Volume II: Implementation and Analysis, Primary and Middle School Years*. Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., and Kelly, D.L. (Eds.). (1999). *TIMSS Technical Report: Volume III: Implementation and Analysis, Final Year of Secondary School*. Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., Gregory, G.D., and Stemler, S.E. (Eds.). (2000). *TIMSS 1999 Technical Report.* Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., Mullis, I.V.S., and Chrostowski, S.J. (Eds.). (2004). *TIMSS 2003 Technical Report: Findings From IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Boston College, International Study Center. Chestnut Hill, MA.

Olson, J.F., Martin, M.O., and Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 Technical Report.* Boston College International Study Center, Chestnut Hill.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Data Quality and Comparability

Martin, M.O., and Mullis, I.V.S. (Eds.). (1996). *Quality Assurance in Data Collection.* Boston College, International Study Center. Chestnut Hill, MA.

Chapter 22: Program for International Student Assessment (PISA)

1. OVERVIEW

INTERNATIONAL ASSESSMENT OF 15-YEAR-OLDS:

Assesses literacy skills in the following areas:

- Reading literacy
- Mathematics literacy
- Science literacy

The Program for International Student Assessment (PISA) is a system of international assessments that measures 15-year-old students' capabilities in reading literacy, mathematics literacy, and science literacy every 3 years. PISA 2006 was the third in this series of assessments; the fourth in the series took place in 2009. Information on PISA 2009 will not be available until December 2010, so PISA 2009 will not be included in some sections of this chapter. PISA, first implemented in 2000, was developed and is administered under the auspices of the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of industrialized countries.[1] The PISA Consortium, a group of international organizations engaged by the OECD, is responsible for coordinating the study operations across countries. The National Center for Education Statistics (NCES), in the Institute of Education Sciences at the U.S. Department of Education, is responsible for the implementation of PISA in the United States.

Purpose

PISA provides internationally comparative information on the reading, mathematics, and science literacy of students at an age that, for most jurisdictions, is near the end of compulsory schooling. The objective of PISA is to measure the "yield" of education systems, or what skills and competencies students have acquired and can apply in reading, mathematics, and science to real-world contexts by age 15. The literacy concept emphasizes the mastery of processes, the understanding of concepts, and the application of knowledge and functioning in various situations. By focusing on literacy, PISA draws not only from school curricula but also from learning that may occur outside of school.

Components

Assessment. PISA is a paper-and-pencil assessment that is designed to assess 15-year-olds' performance in reading, mathematics, and science literacy. Each student takes a 2-hour assessment. Assessment items include a combination of multiple-choice questions, closed- or short- response questions (for which answers are either correct or incorrect), and open-constructed response questions (for which answers can receive partial credit). PISA scores are reported on a scale of 0 to 1,000 with a scale mean of 500 and a scale standard deviation of 100.

Questionnaires. Students complete a 30-minute questionnaire providing information about their backgrounds, attitudes, and experiences in school. In addition, the principal of each participating school completes a 20- to 30-minute questionnaire on school characteristics and policies.

[1] Countries that participate in PISA are referred to as jurisdictions throughout this chapter.

Periodicity

PISA operates on a 3-year cycle. Each PISA assessment cycle focuses on one subject in particular, although all three subjects are assessed every year. In PISA 2000, reading literacy was the major focus. In 2003, PISA focused on mathematics literacy, and in 2006, PISA focused on science literacy. In 2009, PISA again focused on reading literacy. The remainder of this chapter focuses on the design of the 2006 administration.

2. USES OF DATA

PISA provides valuable information for comparisons of student performance across jurisdictions and over time at the national level and for some countries the subnational level. Performance in each subject area can be compared across jurisdictions in terms of:

- Jurisdictions' mean scores;
- The proportion of students in each jurisdiction reaching PISA proficiency levels;
- The scores of jurisdictions' highest performing and lowest performing students;
- The standard deviation of the distribution of scores in each jurisdiction; and
- Other measures of the distribution of performance within jurisdictions.

PISA also supports cross-jurisdictional comparisons of the performance of some subgroups of students, including students grouped by sex, immigrant status, and socioeconomic status. PISA data are not useful for comparing the performance of racial/ethnic groups across jurisdictions, because relevant racial/ethnic groups differ across jurisdictions. However, U.S. PISA datasets include information that can be used in comparing groups of students by race/ethnicity, and the poverty level of their schools within the country.

Contextual measures taken from student and principal questionnaires can be used to compare the educational contexts of 15-year-old students across jurisdictions. Caution should be taken in attempting to interpret associations between measures of educational context and student performance. The PISA assessment is intended to tap the knowledge and skills developed by students over several years as they develop factual knowledge and problem-solving skills and learn to apply them in a variety of situations. PISA contextual measures typically refer to students' current school context, which may differ from their prior school context. In the United States, data collection occurs in the fall of the school year; therefore, contextual measures may apply to only 1 or 2 months of school.

Through the collection of comparable information across jurisdictions at the student and school levels, PISA adds significantly to the knowledge base that was previously available from official national statistics.

3. KEY CONCEPTS

Literacy Types

The types of literacy measured by PISA are defined as follows (OECD 2009).

Reading literacy. An individual's capacity to understand, use, reflect on and engage with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.

Mathematics literacy. An individual's capacity to identify and understand the role that mathematics plays in the world, make well-founded judgments, and use and engage with mathematics in ways that meet one's needs as a constructive, concerned, and reflective citizen.

Science literacy. An individual's scientific knowledge and the use of that knowledge to identify questions, acquire new knowledge, explain scientific phenomena, and draw evidence-based conclusions about science-related issues; an understanding of the characteristic features of science as a form of human knowledge and inquiry; an awareness of how science and technology shape our material, intellectual, and cultural environments; and a willingness to engage in science-related issues—and with the ideas of science—as a reflective citizen.

4. SURVEY DESIGN

The survey design for the PISA 2006 data collection is discussed in this section.

Target Population

Each jurisdiction was required to follow international standards for designing and selecting the sample, as given in the PISA sampling manual for the 2006 assessment (PISA Project Consortium 2005b). The international sampling guidelines defined the target

population and set the requirement for participation rates. The *desired* national PISA target population consisted of 15-year-old students attending educational institutions located within the jurisdiction, in 7th grade and higher. Jurisdictions were to include 15-year-old students enrolled full time in educational institutions, enrolled part time in educational institutions, enrolled in a vocational training or related type of educational program, and attending a foreign school within the jurisdiction (as well as students from other jurisdictions attending any of the programs in the first three categories). It was recognized that no testing of persons schooled in the home, workplace, or out of the jurisdiction occurred; therefore, these students were not included in the international target population.

The operational definition of an age population directly depends on the testing dates. International standards required that students in the sample be 15 years and 3 months to 16 years and 2 months at the beginning of the testing period. For PISA 2006, the testing period suggested by the OECD was between March 1, 2006, and August 31, 2006, and was required not to exceed 42 days. The United States was one of three jurisdictions to administer the assessment in fall 2006, rather than spring 2006. The United States made this choice to avoid conflicting with mandatory high-stakes testing that often occurs in the spring, based upon the PISA 2003 experience. The United States, the United Kingdom (except for Scotland), and Bulgaria moved their test date to the fall; consequently, the range of eligible birthdates in these jurisdictions was adjusted to ensure that the mean age remained consistent across all jurisdictions. In the United States, students born between July 1, 1990, and June 30, 1991, were eligible to participate in PISA 2006.

International Sample Design

In the 2006 PISA assessment, most jurisdictions used a two-stage stratified sample. The first-stage sampling units consisted of individual schools having 15-year-old students. In all but a few jurisdictions, schools were sampled systematically from a comprehensive national list of all eligible schools with probabilities that were proportional to a measure of size. This is referred to as probability proportional to size (PPS) sampling. The measure of size was a function of the estimated number of eligible 15-year-old students enrolled in the school. Prior to sampling, schools in the sampling frame were assigned to strata formed either explicitly or implicitly. The second-stage sampling units in jurisdictions using the two-stage design consisted of students within sampled schools. Once a school was selected to be in the sample, a list of the school's 15-year-old students was prepared. From each list that contained more than 35 students, 35 students were selected with equal probability, and for lists of fewer than 35 students, all students were selected. However, the minimum number of students that could be sampled within a school was 20.

Because PISA is an international survey, the types of exclusions must be defined internationally and the exclusion rates have to be limited in order to ensure that survey results are representative of the entire national school system. Thus, efforts were made to guarantee that exclusions, if they were necessary, were minimized. Exclusions could take place at the school selection stage (by excluding the whole school) or at the student selection stage.

International within-school exclusion rules for students were specified as follows:

- *Students with functional disabilities.* These were students with a moderate to severe permanent physical disability such that they could not perform in the PISA testing environment.

- *Students with intellectual disabilities.* These were students with a mental or emotional disability who had been tested as cognitively delayed or who were considered in the professional opinion of qualified staff to be cognitively delayed such that they could not perform in the PISA testing situation.

- *Students with insufficient language experience.* These were students who met the three criteria of (1) not being a native speaker in the assessment language, (2) having limited proficiency in the assessment language, and (3) having received less than a year of instruction in the assessment language. In the United States, English was the exclusive language of the assessment.

A school attended only by students who would be excluded for intellectual, functional, or linguistic reasons was considered a school-level exclusion.
School-level exclusions for inaccessibility, feasibility, or other reasons were required to cover fewer than 0.5 percent of the total number of students in the international PISA target population. International guidelines state that no more than 5 percent of a jurisdiction's desired national target population should be excluded from the sample.

A minimum of 150 schools (or all schools, if there were fewer than 150 in a participating jurisdiction) had to be selected in each jurisdiction. Within each participating school, a sample of the PISA-eligible

students was selected with equal probability. In total, a minimum sample size of 4,500 assessed students was to be achieved. If a jurisdiction had fewer than 4,500 eligible students, then the sample size was the *national defined target population*. The national defined target population included all those eligible students in the schools that were listed in the school sampling frame.

Response Rate Targets

School response rates. The PISA international guidelines for the 2006 assessment required that jurisdictions achieve an 85 percent school response rate. However, while stating that each jurisdiction must make every effort to obtain cooperation from the sampled schools, the requirements also recognized that this is not always possible. Thus, it was allowable to use substitute, or replacement, schools as a means to avoid loss of sample size associated with school nonresponse. The international guidelines stated that at least 65 percent of participating schools must be from the original sample. Jurisdictions were allowed to use replacement schools (selected during the sampling process) to increase the response rate once the 65 percent benchmark had been reached.

Each sampled school was to be assigned two replacement schools in the sampling frame. If the original sample school refused to participate, a replacement school was asked to participate. The international guidelines define the response rate as the number of participating schools (both original and replacement schools) divided by the total number of eligible original sample schools.[2]

Student response rates. A minimum response rate of 80 percent of selected students across participating schools was required. Students were deemed participants if they gave at least one response to the cognitive assessment, or if they responded to at least one student questionnaire item and either they or their parents provided the occupation of a parent or guardian.

Within each school, a student response rate of 50 percent was required for a school to be regarded as participating: the overall student response rate was computed using only students from schools with at least a 50 percent response rate. Weighted student response rates were used for assessing this standard. Each student was weighted by the reciprocal of his or her sample selection probability.

Sample Design in the United States

The design of the U.S. school sample for PISA 2006 was developed to achieve each of the international requirements set forth in the PISA sampling manual. The U.S. school sample is self-weighting, is stratified, consists of two stages (described below), and was selected using PPS sampling. The measure of size used in the first stage was the expected number of eligible 15-year-old students in the school. At the second stage, a sample of 42 students was selected from each school, regardless of size (all eligible students were selected if there were fewer than 42).

A list of schools for the U.S. sample was prepared using data from the 2003–04 Common Core of Data (CCD) and the 2003–04 Private School Universe Survey (PSS), two NCES surveys. These schools were stratified into two explicit groups: schools with large enrollments of 15-year-old students and schools with small enrollments of 15-year-old students. The frame was implicitly stratified (i.e., sorted for sampling) by five categorical stratification variables: grade span of school; control of school (public or private); region of the country; type of location relative to populous areas; and percentage of students of Black, Hispanic, and other race/ethnicities (above or below 15 percent). The last variable used for sorting within the implicit stratification was the estimated enrollment of 15-year-olds based on grade enrollments.

As in PISA 2003, schools were selected in the first stage with PPS, and students were sampled in the second stage, yielding overall equal probabilities of selection. In PISA 2000, the U.S. school sample had a three-stage design, the first of which was the selection of a sample of geographic primary sampling units (PSUs). The change to a two-stage model was made in PISA 2003 to reduce the design effects observed in the 2000 data and to minimize respondent burden on individual districts by spreading it across school districts as much as possible.

Once the school sample was drawn, it was loaded into KeyQuest, a software program written specifically for jurisdictions participating in PISA. KeyQuest was used to manage the sample, draw the student sample, track participation, and produce verification reports used to clean the data in preparation for submitting the data file.

The U.S. school sample for PISA 2006 consisted of 240 schools (from 44 states) containing at least one 7th through 12th grade class. There were 27 sampled

[2] The calculation of response rates described here is based on the formula stated in the international guidelines and is not consistent with NCES standards. A more conservative way to calculate response rates would be to include participating replacement schools in the denominator as well as in the numerator and to add replacement schools that were hard refusals to the denominator.

schools identified as ineligible or closed, reducing the sample to 209 schools.

Assessment Design

Test scope and format. In PISA 2006, the three subject domains were tested, with science as the major domain and reading and mathematics as minor domains. The development of the PISA 2006 assessment instruments was an interactive process among the PISA Project Consortium, various expert committees, and OECD members. The assessment included items submitted by participating jurisdictions and items developed by the consortium's test developers. Representatives of each jurisdiction reviewed the items for possible bias and for relevance to PISA's goals. The intention was to reflect in the assessment the national, cultural, and linguistic variety of the OECD jurisdictions. Science items were field tested in 2005 in each jurisdiction to examine their psychometric properties and identify any problematic items. Mathematics and reading items were field tested in jurisdictions that had not participated in PISA 2003. Following the field test, statistics were reviewed for each item for each jurisdiction, including percent correct, item difficulty, item discrimination, and gender differences. Items that worked differently across jurisdictions were deleted.

PISA 2006 was a paper-and-pencil assessment. Approximately one-third of the science literacy items were multiple-choice items, one-third were closed- or short-response items (for which students wrote an answer that was simply either correct or incorrect), and about one-third were open constructed-response items (for which students wrote answers that could be assigned partial credit). Items other than multiple choice were graded by trained scorers using an international scoring guide specific to each item that explicated the requirements for each score level.

Multiple-choice items were either (a) standard multiple choice, with a limited number (usually four) of responses from which students were required to select the best answer; or (b) complex multiple choice, which presented several statements, each of which required students to choose one of several possible responses (true/false, correct/incorrect, etc.). Closed- or short-response items included items which generally required students to construct a response within very limited constraints, such as mathematics items requiring a numeric answer, and items requiring a word or short phrase. Open constructed-response items required more extensive writing, or showing a calculation, and frequently included some explanation or justification. Pencils, erasers, rulers, and in some cases, calculators were provided.

Test design. The final 2006 assessment consisted of 140 science items, 48 mathematics items, and 28 reading items.

In order to cover the intended broad range of content while meeting the limit of 2 hours of individual testing time, the assessment in each domain was divided into clusters and organized into 13 booklets. Each booklet was made up of four test clusters. There were seven science clusters, four mathematics clusters, and two reading clusters. The clusters were allocated in a rotated design to the 13 booklets. The average number of items per cluster was 20 for science, 12 for mathematics, and 14 for reading. Each cluster was designed to average 30 minutes of test material.

The sampled students were randomly assigned one of the booklets, which meant each student undertook 2 hours of testing. The 2-hour test booklets were arranged in two 1-hour parts, each made up of two 30-minute time blocks. PISA's procedures provided for a short break to be taken between administrations of the two parts of the test booklet.

Every student answered science items, while mathematics and reading items were spread throughout the booklets. This assessment design was balanced so that each item cluster appeared four times, once in each of four possible locations in a booklet. Furthermore, each cluster appeared once with each other cluster. The final design, therefore, ensured that a representative sample of students responded to each cluster of items. The linked design enabled standard measurement techniques to be applied to the resulting student response data to estimate item difficulties and student abilities.

In addition to the cognitive assessment, students also received a 30-minute questionnaire designed to elicit information about their backgrounds, their attitudes, and their experiences in school. Principals in schools where PISA was administered also received a 20- to 30-minute questionnaire about their schools.

In addition to the 13 two-hour booklets, a special, optional one-hour booklet, referred to as the UH booklet (or the Une Heure booklet), was prepared for use in schools catering exclusively to students with special needs. The United States did not use the optional one-hour test booklet.

Test printing. The data collection contractor for PISA 2006, RTI International, made an error printing the test booklets in the United States, and the pagination of the booklets was consistently off by one page. The international consortium intended for the first page to

be printed on the inside of the cover; in the United States it was typically printed on the first page of plain white paper. As a result, some of the instructions in the reading section were incorrect. In some passages, students were incorrectly instructed to refer to the passage on the "opposite page" when the passage now appeared on the previous page. Because of the small number of items in the reading section, it was not possible to recalibrate the score to exclude the affected items. No incorrect page references appeared in the mathematics or science sections of the booklets.[3]

Data Collection and Processing

PISA is implemented in each jurisdiction by a National Project Manager (NPM). In the United States, the NPM works with a national data collection contractor to implement procedures prepared by the International Consortium and agreed to by the participating jurisdictions. In 2006, the U.S. national data collection contractor was RTI International.

Reference dates. The testing period suggested by the OECD was between March 1, 2006, and August 31, 2006, and was required not to exceed 42 days. However, the United States, in order to improve response rates and better accommodate school schedules, scheduled the PISA 2006 data collection from September 25 to November 22, 2006, with the agreement of the PISA Consortium. The United Kingdom (except Scotland) and Bulgaria also opted for a fall data collection period for PISA 2006.

Data collection. To implement the PISA 2006 assessment in schools, the NPMs were assisted by school coordinators and test administrators. Each NPM typically had several assistants, working from a base location (referred to as a national center). The NPM manual provided detailed information about the duties and responsibilities of the NPM. Supplementary manuals, with detailed information about particular aspects of the project, were also provided.

The test administrators were primarily responsible for administering the PISA 2006 test fairly, impartially, and uniformly, in accordance with international standards and PISA 2006 procedures. To maintain fairness, international standards stipulated that test administrators could not be the reading, mathematics, or science teacher of the students being assessed, and it was preferred that they not be a staff member at any participating school. Prior to the test date, test administrators were trained by national centers. Training included a thorough review of the test administrator manual and the script to be followed during the administration of the test and questionnaire. The PISA Project Consortium prepared a test administrator manual that described in detail the activities and responsibilities of the test administrator.

Four field supervisors and 35 test administrators were hired to work on the PISA 2006 main study in the United States. Each test administrator was assigned to one of the four field supervisors, who coordinated and monitored the test administrator's work.

The test administrator followed the instructions set forth in the international PISA test administrator manual. The students were randomly assigned one of 13 test booklets. Test administrators distributed the assessment booklets, matching the student with the preassigned booklet type according to the preprinted Student Tracking Form.

The principal at each participating school designated one person to serve as the school coordinator for PISA 2006. School coordinators were asked to work with project staff to coordinate the logistics of the test session and to ensure high student response rates. School coordinators coordinated school-related activities with the national center and the test administrators. On the test day, the school coordinator was expected to ensure that the sampled students attended the test session(s). If necessary, the school coordinator also made arrangements for a follow-up session and ensured that absent students attended the follow-up session. The PISA Project Consortium prepared a school coordinator manual that described in detail the activities and responsibilities of the school coordinator.

In the United States, schools were offered the option of conducting the assessment after school hours or on a Saturday, in addition to during school hours. This option was offered only as a refusal conversion tool and not as part of the initial recruitment materials. Of the 166 participating schools, 88 schools conducted the session during school hours, 4 conducted the session after school, and 74 participated on a Saturday. The student response rate was 91 percent during school hours and 90 percent in schools where PISA 2006 was administered after school or on a Saturday. Analyses were conducted comparing the performance of students who took the test during the regular school day with those who took the exam after school or on a Saturday. No measurable differences were found between the two groups.

Scoring. At least one-third of the PISA 2006 assessment was devoted to open constructed responses.

[3] Because of this printing error, the OECD and NCES decided not to report the PISA 2006 reading results for the United States.

The process of scoring these items was an important step in ensuring the quality and comparability of the PISA 2006 data. Detailed guidelines were developed for the scoring guides themselves, training materials to recruit scorers, and workshop materials used for the training of national scorers. Prior to the national training, the PISA Project Consortium organized training sessions to present the material and train the scoring coordinators from the participating jurisdictions, who in turn trained the national scorers.

For each test item, the scoring guide described the intent of the question and how to code the students' responses to each item. This description included the credit labels—full credit, partial credit, or no credit—attached to the possible categories of response. Also included was a system of double-digit coding for some mathematics and science items, where the first digit represented the score and the second digit represented different strategies or approaches that students used to solve the problem. The second digit generated national profiles of student strategies and misconceptions. In addition, the scoring guides included real examples of students' responses accompanied by a rationale for their classification for purposes of clarity and illustration.

To examine the consistency of this marking process in more detail within each jurisdiction (and to estimate the magnitude of the variance components associated with the use of scorers), the PISA Project Consortium conducted an interscorer reliability study on a subsample of assessment booklets. Homogeneity analysis was applied to the national sets of multiple scoring and compared with the results of the field trial. A full description of this process and the results can be found in the OECD's *PISA 2006 Technical Report* (OECD 2009).

Data Entry and Verification. The PISA Project Consortium provided participating jurisdictions with the KeyQuest data entry software. KeyQuest contained the database structures for all of the booklets, questionnaires, and tracking forms used in the main survey. Variables could be added or deleted as needed for different national options. Approved adaptations to response categories could also be accommodated. Student response data were entered directly from the test booklets and questionnaires. NPMs were responsible for ensuring that their jurisdiction's data underwent many quality checks before the data files were submitted to the PISA Project Consortium.

Once the data files were submitted to the PISA Project Consortium, they underwent two independent data cleaning procedures by data analysts. During cleaning, as many anomalies and inconsistencies as possible were identified, and through a process of extensive discussion between each national center and the PISA Project Consortium's data processing center at the Australian Council for Educational Research (ACER), an effort was made to correct and resolve all data issues.

Estimation Methods

Weighting. The use of sampling weights is necessary for the computation of statistically sound, nationally representative estimates. Survey weights adjust for the probabilities of selection for individual schools and students, for school or student nonresponse, and for errors in estimating the size of the school or the number of 15-year-olds in the school at the time of sampling.

The internationally defined weighting specifications for PISA 2006 included two base weights and five adjustments. The school base weight was defined as the reciprocal of the school's probability of selection. (For replacement schools, the school base weight was set equal to the weight of the original school it replaced.) The student base weight was given as the reciprocal of the probability of selection for each selected student from within a school.

These base weights were then adjusted for school and student nonresponse. The school nonresponse adjustment was done individually for each jurisdiction using implicit and explicit strata defined as part of the sample design. In the case of the United States, three variables were used: school control, census region, and community type. The student nonresponse adjustment was done within cells based first on students' final school nonresponse cell and their explicit stratum; within that, grade and gender were used.

All PISA 2006 analyses were conducted using these sampling weights.

Scaling. There were 13 test booklets, each containing a slightly different subset of items, in the PISA 2006 design. Each student completed one test booklet. The fact that each student completed only a subset of items means that classic test scores, such as the percent correct, are not accurate measures of student performance. Instead, scaling techniques were used to establish a common scale for all students. For PISA 2006, item response theory (IRT) was used to estimate average scores for science, mathematics, and reading literacy for each jurisdiction.

IRT identifies patterns of response and uses statistical models to predict the probability of a student

answering an item correctly as a function of his or her proficiency in answering other questions. PISA 2006 used a mixed coefficients multinomial logit IRT model. This model is similar in principle to the more familiar two-parameter logistic IRT model. With this method, the performance of a sample of students in a subject area or subarea can be summarized on a simple scale or series of scales, even when students are administered different items.

Plausible values. Scores for students are estimated as plausible values because each student completed only a subset of items. These values represent the distribution of potential scores for all students in the population with similar characteristics and identical patterns of item response. It is important to recognize that plausible values are not test scores and should not be treated as such. Plausible values are randomly drawn from the distribution of scores that could be reasonably assigned to each individual. As such, the plausible values contain random error variance components and are not optimal as scores for individuals.

The PISA 2006 student file contains many plausible values, five for each of the PISA 2006 cognitive scales (combined science literacy scale, three science literacy subscales, reading literacy scale, and mathematics literacy scale). If an analysis is to be undertaken with one of these cognitive scales, then (ideally) the analysis should be undertaken five times, once with each of the five relevant plausible value variables. The results of these five analyses are averaged; then, significance tests that adjust for variation between the five sets of results are computed.

Imputation. As with all item response scaling models, student proficiencies (or measures) are not observed; they are missing data that must be inferred from the observed item responses. There are several possible alternative approaches for making this inference. PISA uses the imputation methodology usually referred to as plausible values (described above). Plausible values are a selection of likely proficiencies for students that attained each score. Missing background data from student and principal questionnaires are not imputed for PISA reports published by NCES. In general, item response rates for variables discussed in NCES PISA reports are over the NCES standard of 85 percent.

Measuring trends. Reading literacy scales used in PISA 2000, 2003, and 2006 are directly comparable, which means that the value of 500 in PISA 2006 has the same meaning as it did in PISA 2000 and PISA 2003. However, since mathematics literacy was the major domain assessed in PISA 2003, the mathematics assessment underwent major development work and was broadened to include four domains; only two of these domains appeared in PISA 2000. As such, mathematics literacy scales are only comparable between PISA 2003 and PISA 2006. Likewise, PISA 2006 was the first major assessment of science literacy. As such, the science literacy scale in PISA 2006 is not directly comparable with those of earlier PISA assessments; however, it establishes the basis for monitoring future trends in science performance.

The PISA 2000, PISA 2003, and PISA 2006 assessments of reading, mathematics, and science are linked assessments. That is, the sets of items used to assess each domain in each year include a subset of common items. Between PISA 2000 and PISA 2003, there were 28 reading items (units and clusters), 20 mathematics items, and 25 science items that were used in both assessments. These common items are referred to as link items. The same 28 reading items were retained in 2006 to link the PISA 2006 data to PISA 2003, while 48 mathematics items from PISA 2003 were used in PISA 2006. For the science assessment, just 22 items were common to PISA 2006 and PISA 2003, and 14 were common to PISA 2006 and PISA 2000.

To establish common reporting metrics for PISA, the difficulty of the link items, measured on different occasions, is compared. Using procedures that are detailed in the *PISA 2006 Technical Report* (OECD 2009), the comparison of the item difficulties on the different occasions is used to determine a score transformation that allows the reporting of the data for a particular subject on a common scale. The change in the difficulty of each of the individual link items is used in determining the transformation; as a consequence, the sample of link items that has been chosen will influence the choice of transformation. This means that if an alternative set of link items had been chosen, the resulting transformation would be slightly different. The consequence is an uncertainty in the transformation due to the sampling of the link items, just as there is an uncertainty in values such as jurisdiction means due to the use of a sample of students.

Future Plans

After the release of PISA 2009 results in December of 2010, the next PISA assessment will be conducted in 2012. The major domain in PISA 2012 will be mathematics literacy. PISA 2012 will also include, in addition to paper-and-pencil assessments in mathematics, science, and reading literacy, computer-based assessments in mathematics and reading and a computer-based problem-solving assessment.

5. DATA QUALITY AND COMPARABILITY

A comprehensive program of continuous quality monitoring was central to ensuring full, valid implementation of the PISA 2006 procedures and the recording of deviations from these procedures. Quality monitors from the PISA Consortium visited a sample of schools in every jurisdiction to ensure that testing procedures were carried out in a consistent manner. The purpose of quality monitoring is to observe and record the implementation of the described procedures; therefore, the field operations manuals provided the foundation for all the quality monitoring procedures.

The manuals that formed the basis for the quality monitoring procedures were the NPM manual, the test administrator manual, the school coordinator manual, the school sampling preparation manual, and the PISA data management manual. In addition, the PISA data were verified at several points starting at the time of data entry.

Despite the efforts taken to minimize error, as with any study, PISA has limitations that researchers should take into consideration. Below are discussed two possible sources of error in PISA, sampling and nonsampling errors.

Sampling Error

Sampling errors occur when a discrepancy between a population characteristic and the sample estimate arises because not all members of the target population are sampled for the survey. The size of the sample relative to the population and the variability of the population characteristics both influence the magnitude of sampling error. The particular sample of 15-year-old students from fall 2006 was just one of many possible samples that could have been selected. Therefore, estimates produced from the PISA 2006 sample may differ from estimates that would have been produced had another sample of students been selected. This type of variability is called sampling error because it arises from using a sample of 15-year-old students in 2006 rather than all 15-year-old students in that year.

The standard error is a measure of the variability owing to sampling when estimating a statistic. The approach used for calculating sampling variances in PISA is Balanced Repeated Replication (BRR). This method of producing standard errors uses information about the sample design to produce more accurate standard errors than would be produced using simple random sample assumptions. Thus, the standard errors reported in PISA can be used as a measure of the precision expected from this particular sample.

Nonsampling Error

“Nonsampling error” is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. For example, the sampling frame in the United States was limited to regular public and private schools in the 50 states and the District of Columbia and cannot be used to represent Puerto Rico or other jurisdictions. The sources of nonsampling errors are typically problems such as unit and item nonresponse, the differences in respondents’ interpretations of the meaning of survey questions, response differences related to the particular time the survey was conducted, and mistakes in data preparation.

In general, it is difficult to identify and estimate either the amount of nonsampling error or the bias caused by this error. In PISA 2006, efforts were made to prevent such errors from occurring and to compensate for them when possible. For example, the design phase entailed a field test that evaluated items as well as the implementation procedures for the survey. Another potential source of nonsampling error was respondent bias, which occurs when respondents systematically misreport (intentionally or unintentionally) information in a study. One potential source of respondent bias in this survey was social desirability bias. For example, students may overstate their parents’ educational attainment or occupational status. If there were no systematic differences among specific groups under study in their tendency to give socially desirable responses, then comparisons of the different groups would accurately reflect differences among groups. Readers should be aware that respondent bias may be present in this survey as in any survey. It was not possible to state precisely how such bias may affect the results.

Coverage error. Every NPM was required to define and describe their jurisdiction’s national desired target population and explain how and why it might deviate from the international target population. Any hardships in accomplishing complete coverage were specified, discussed, and approved (or not) in advance. Where the national desired target population deviated from full national coverage of all eligible students, the deviations were described and enrollment data provided to measure how much that coverage was reduced. School-level and within-school exclusions from the national desired target population resulted in a national defined target population corresponding to the population of

students recorded in each jurisdiction's school sampling frame.

In PISA 2006, the United States reported 85 percent coverage of the 15-year-old population and 96 percent coverage of the national desired population. The United States reported a 4.3 percent exclusion rate, which was below the internationally acceptable exclusion rate of 5 percent.

Nonresponse error. Nonresponse error results from nonparticipation of schools and students. School nonresponse, where no replacement school participated in PISA 2006, will lead to the underrepresentation of students from the type of school that did not participate, unless weighting adjustments are made. It is also possible that only a part of the eligible population in a school (such as those 15-year-olds in a single grade) was represented by the school's student sample; this also requires weighting to compensate for the missing data from the omitted grades. Student nonresponse within participating schools occurred to varying extents. Students that could not be given achievement test scores (described in more detail below), but were not excluded for linguistic or disability reasons will be underrepresented in the data unless weighting adjustments are made.

Unit Nonresponse. In PISA 2006 in the United States, of the 240 sampled schools, 210 were eligible and 150 agreed to participate. The school response rate before replacement was 69 percent (weighted and unweighted) (Table 16). In addition to the 150 participating original schools, 20 replacement schools participated, for a total of 170 participating schools and a school response rate of 79 percent (weighted and unweighted).[4] Each of the participating schools achieved over 50 percent student participation and was included in the overall student response rate calculations.

A total of 6,800 students in the United States were sampled for the assessment. Of these students, 37 were deemed ineligible because of their enrolled grades or birthdays and 330 were deemed ineligible because they had left the school. These students were removed from the sample. Of the eligible 6,430 sampled students, an additional 250 were excluded using the criteria described earlier, for a weighted exclusion rate of 3.8 percent at the student level. Combined with the 0.5 percent of students excluded at the school level, before sampling, the overall exclusion rate for the United States was 4.3 percent. Of the 6,180 remaining sampled students, 5,620 participated. During data processing, 10 cases were deleted, leaving 5,610 cases in the final U.S. data file, for a weighted and unweighted student participation rate of 91 percent.

Table 16. U.S. weighted and unweighted school and student response rates: PISA 2006

	Response rate (percent)	
	Weighted	Unweighted
School		
Before replacement	69.0	69.4
After replacement	79.1	79.4
Student	91.0	90.8

SOURCE: Green, P., Herget, D., and Rosen, J. (2009). *User's Guide for the Program for International Student Assessment (PISA): 2006 Data Files and Database With United States Specific Variables* (NCES 2009-055). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

Item nonresponse. For PISA 2006 in the United States, an item nonresponse bias analysis was conducted for the seven school questionnaire items with a response rate less than 85 percent and for the eight student questionnaire items with a response rate less than 85 percent. For each questionnaire item, respondents for that item were compared with nonrespondents for that item based on demographic characteristics known for everyone. These characteristics are from the CCD and PSS files, and continuous variables were made into categorical variables based on quartiles for the purpose of this analysis. For each category of each variable, bias was computed as the percentage of all item respondents who are in that category minus the percentage of all item nonrespondents who are in that category.

In PISA 2006 in the United States, five of the seven questionnaire items were significantly biased for public and private school types. There was no significant bias for any of the categories for the characteristics of total school enrollment, percent White student enrollment, and percent other student enrollment. For more details, refer to *User's Guide for the Program for International Student Assessment (PISA): 2006 Data Files and Database with United States Specific Variables* (Green, Herget, and Rosen 2009).

Measurement error. Measurement error is introduced into a survey when its test instruments do not accurately measure the knowledge or aptitude they are intended to assess.

[4] Since the U.S. school response rate was lower than the international requirement of 85 percent, the PISA Project Consortium required NCES to provide a detailed analysis of school nonresponse bias, which indicated no evidence of substantial bias resulting from school nonresponse (Green, Herget, and Rosen 2009).

Data Comparability

A number of international comparative studies already exist to measure achievement in mathematics, science, and reading, including the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). The Adult Literacy and Lifeskills Survey (ALL) was last conducted in 2003 and measured the literacy and numeracy skills of adults. A new study, the Program for the International Assessment of Adult Competencies (PIAAC), will be administered for the first time in 2011 and will assess the level and distribution of adult skills required for successful participation in the economy of participating jurisdictions. In addition, the United States has been conducting its own national surveys of student achievement for more than 35 years through the National Assessment of Educational Progress (NAEP). PISA differs from these studies in several ways.

Content. PISA is designed to measure "literacy" broadly, whereas studies such as TIMSS and NAEP have a stronger link to curriculum frameworks and seek to measure students' mastery of specific knowledge, skills, and concepts. The content of PISA is drawn from broad content areas (e.g., space and shape in mathematics) in contrast to more specific curriculum-based content, such as geometry or algebra.

Tasks. PISA also differs from other assessments in that it emphasizes the application of reading, mathematics, and science literacy to everyday situations by asking students to perform tasks that involve interpretation of real-world materials as much as possible. A study comparing the PISA, NAEP, and TIMSS mathematics assessments (Neidorf et al. 2006) found that the mathematics topics addressed by each assessment are similar, although PISA places greater emphasis on data analysis and less on algebra than does either NAEP or TIMSS. However, it is in how that content is presented that makes PISA different.

PISA uses multiple-choice items to a far lesser degree than NAEP or TIMSS, and it contains a higher proportion of items reflecting moderate to high mathematical complexity than do those two assessments. An earlier comparative analysis of the PISA, TIMSS, and NAEP mathematics and science assessments (Nohara 2001) found that compared with NAEP and TIMSS, more items in the PISA science assessment built connections to practical situations and required students to demonstrate multi-step reasoning, and fewer items used a multiple-choice format. The study also found that compared with NAEP and TIMSS, more items in the PISA mathematics assessment were set in real-life situations or scenarios, required multi-step reasoning, and required interpretation of figures and other graphical data. These tasks reflect the underlying assumption of PISA: as 15-year-olds begin to make the transition to adult life, they not only need to know how to read or use particular mathematical formulas or scientific concepts, but they also need to know how to apply this knowledge and these skills in the many different situations they will encounter in their lives.

Age-based sample. In contrast with TIMSS and PIRLS, which are grade-based assessments, PISA's sample is based on age. TIMSS assesses fourth- and eighth-graders, while PIRLS assesses fourth-graders only. The PISA sample, however, is drawn from 15-year-old students, regardless of grade level. The goal of PISA is to represent outcomes of learning rather than outcomes of schooling. By placing the emphasis on age, PISA intends to show not only what 15-year-olds have learned in school in a particular grade, but outside of school as well as over the years. PISA thus seeks to show the overall yield of an education system and the cumulative effects of all learning experience. Focusing on age 15 provides an opportunity to measure broad learning outcomes while all students are still required to be in school across the many participating jurisdictions. Finally, because years of education vary among jurisdictions, choosing an age-based sample makes comparisons across jurisdictions somewhat easier.

6. CONTACT INFORMATION

For content information about the PISA, contact:

Daniel McGrath
Phone: (202) 502-7426
E-mail: daniel.mcgrath@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

Most of the technical documentation for PISA is published by the OECD. The U.S. Department of

Education, NCES, is the source of several additional references listed below.

General

Baldi, S., Jin, Y., Skemer, M., Green, P., Herget, D., and Xie, H. (2007). *Highlights From PISA 2006: Performance of U.S. 15-Year-Olds in Science and Mathematics Literacy in an International Context* (NCES 2008-016). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., and Jocelyn, L. (2004). *International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results From the U.S. Perspective* (NCES 2005-003). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Organization for Economic Cooperation and Development (OECD). (2005). *PISA 2003 Data Analysis Manual*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2009). *PISA 2009 Assessment Framework - Key Competencies in Reading, Mathematics and Science*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis.* Paris: Author.

Survey Design

Green, P., Herget, D., and Rosen, J. (2009). *User's Guide for the Program for International Student Assessment (PISA): 2006 Data Files and Database With United States Specific Variables* (NCES 2009-055). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Organization for Economic Cooperation and Development (OECD). (2009). *PISA 2006 Technical Report*. Paris: Author.

PISA Project Consortium. (2004). *Technical Standards for PISA 2006.* Retrieved November 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_technical_standards.pdf

PISA Project Consortium. (2005a). *Main Study National Project Manager's Manual.* Retrieved November 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2003_national_project_manager_manual.pdf

PISA Project Consortium. (2005b). *School Sampling Preparation Manual: PISA 2006 Main Study.* Retrieved November 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2003_school_sampling_manual.pdf

PISA Project Consortium. (2005c). *PISA 2006 Main Study Test Administrator's Manual.* Retrieved November 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_test_administrator_manual.pdf

PISA Project Consortium. (2005d). *PISA 2006 Main Study School Coordinator's Manual.* Retrieved November 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_school_coordinators_manual.pdf

Data Quality and Comparability

Dossey, J.A., McCrone, S.S., and O'Sullivan, C. (2006). *Problem Solving in the PISA and TIMSS 2003 Assessments* (NCES 2007-049). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Neidorf, T.S., Binkley, M., Gattis, K., and Nohara, D. (2006). *Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments* (NCES 2006-029). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Nohara, D. (2001). *A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)* (NCES 2001-07). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Stephens, M., and Coleman, M.M. (2007). *Comparing PIRLS and PISA With NAEP in Reading, Mathematics, and Science*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved November 14, 2008, from

http://nces.ed.gov/Surveys/PISA/pdf/comppaper12
082004.pdf

Chapter 23: International Adult Literacy Survey (IALS)

1. OVERVIEW

The 1994 International Adult Literacy Survey (IALS) represented a first attempt to assess the literacy skills of entire adult populations in a framework that provided data comparable across cultures and languages. This collaborative project was designed to inform both education and labor market policy and program development activities in participating countries. The international portion of the study was carried out under the auspices of an International Steering Committee chaired by Canada, with each participating country holding a seat on the committee along with representatives from the Organization for Economic Cooperation and Development (OECD), European communities, and the United Nations Educational, Scientific and Cultural Organization.

1994 INTERNATIONAL STUDY OF ADULT LITERACY

IALS collected:

- Background assessments
- Literacy assessments

In the United States, IALS is the fourth assessment of adult literacy funded by the federal government and conducted by the Educational Testing Service (ETS). The three previous efforts were (1) the 1992 National Adult Literacy Survey (see chapter 19); (2) the Department of Labor's (DOL) 1990 Workplace Literacy Survey; and (3) the 1985 Young Adult Literacy Assessment (funded as an adjunct to the National Assessment of Educational Progress—see chapter 18). In order to maximize the comparability of estimates across countries, IALS chose to adopt the National Adult Literacy Survey methodology and scales. Literacy was defined along three dimensions—prose, document, and quantitative. These were designed to capture an ordered set of information-processing skills and strategies that adults use to accomplish a diverse range of literacy tasks encountered in everyday life. The background data collected in IALS provide a context for understanding the ways in which various characteristics are associated with demonstrated literacy skills.

IALS was originally conducted in eight countries (Canada, Germany, Ireland, the Netherlands, Poland, Sweden, French- and German-speaking Switzerland, and the United States). A second phase was subsequently conducted in five additional countries or territories (Australia, Flemish-speaking Belgium, Great Britain, New Zealand, and Northern Ireland), and in a final phase included an additional nine countries. This chapter focuses on the first phase, in which the United States participated.

Purpose

To (1) develop scales that would permit comparisons of the literacy performance of adults (16 and older) with a wide range of abilities; (2) if such an assessment could be created, describe and compare the demonstrated literacy skills of adults in different countries.

Components

Each IALS country was given a set of model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments. IALS instruments consisted of three parts: (1) a background questionnaire, which collected demographic information about respondents; (2) a set of core literacy tasks, which screened out respondents with very limited literacy skills; and (3) a main booklet of literacy tasks, used to calibrate literacy levels.

Background Questionnaire. The background questionnaire collected information on languages spoken or read; parents' educational attainment and employment; labor force experiences—employment status, recent labor force experiences, and occupation; reading and writing at work and looking for work; participation in adult education classes—courses taken, financial support, purpose; reading and writing in daily life (excluding work or school); family literacy—children's reading habits, the household's access to reading materials, hours spent watching television; and household information—total income and sources of income. The background questionnaire was to be administered in about 20 minutes.

Literacy Assessment—Core Literacy Tasks and Main Literacy Tasks. One hundred and fourteen tasks were grouped into three scales and divided into seven blocks (labeled A through G), which in turn were compiled into seven test booklets (numbered 1 through 7). Each booklet contained three blocks of tasks and was designed to take about 45 minutes to complete. Respondents began the cognitive part of the assessment by performing a set of six "core" tasks. Only those who were able to perform at least two of the six core tasks correctly (93 percent of respondents) were given the full assessment.

Periodicity

The first phase of data collection for IALS was conducted during the autumn of 1994 in Canada, Germany, Ireland, the Netherlands, Poland, Sweden, Switzerland (French and German-speaking cantons), and the United States. Data were collected from a second group of countries or territories—Australia, Flemish-speaking Belgium, Great Britain, New Zealand, and Northern Ireland—in 1995–96. Data were collected from a third group of countries in 1997–98. No second administration is planned.

2. USES OF DATA

IALS was designed to inform both educational and labor market policy and program development activities in participating countries. The primary objectives of the study were to

- shed light on the relationship between microeconomic variables—such as individual literacy, educational attainment, labor market participation and employment, and macroeconomic issues—such as competitiveness, growth, and restructuring;

- identify subpopulations that are economically and socially disadvantaged by their literacy skill profiles; and

- establish the comparability of assessments of adult literacy.

IALS data provide comparable information about the activities and outcomes of educational systems and institutions in participating countries. Such data can lead to improvements in accountability and policymaking. These data are relevant to policy formation due to the growing political, economic, and cultural ties between countries.

3. KEY CONCEPTS

Some of the key concepts related to the IALS literacy assessment are described below.

Literacy. The ability to use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.

Prose Literacy. The ability to read and use texts of varying levels of difficulty that are presented in sentence and paragraph form, including editorials, news stories, poems, and fiction.

Document Literacy. The knowledge and skills required to locate and use information contained in formats such as job applications, payroll forms, transportation schedules, maps, tables, and graphics.

Quantitative Literacy. The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a checkbook, calculating a

tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

Literacy Scales. The three scales used to report the results for prose, document, and quantitative literacy. These scales, each ranging from 0 to 500, are based on those established for the Young Adult Literacy Assessment, the DOL's Workplace Literacy Survey, and the National Adult Literacy Survey. The scores on each scale represent degrees of proficiency along that particular dimension of literacy. The scales make it possible not only to summarize the literacy proficiencies of the total population and of various subpopulations, but also to determine the relative difficulty of the literacy tasks administered in IALS.

The literacy tasks administered in IALS varied widely in terms of materials, content, and task requirements, and thus in difficulty. A careful analysis of the range of tasks along each scale provides clear evidence of an ordered set of information-processing skills and strategies along each scale. To capture this ordering, each scale was divided into five levels that reflect this progression of information-processing skills and strategies: Level 1 (0 to 225), Level 2 (226 to 275), Level 3 (276 to 325), Level 4 (326 to 375), and Level 5 (376 to 500). Level 1 comprised those adults who could consistently succeed with Level 1 literacy tasks but not with Level 2 tasks, as well as those who could not consistently succeed with Level 1 tasks and those who were not literate enough to take the test at all. Adults in Levels 2 through 4 were consistently able to succeed with tasks at their level but not with the next more difficult level of tasks. Adults in Level 5 were consistently able to succeed with Level 5 tasks. The use of three parallel literacy scales makes it possible to profile and compare the various types and levels of literacy demonstrated by adults in different countries and by subgroups within those countries.

4. SURVEY DESIGN

Statistics Canada and ETS, a private testing organization in the United States, coordinated the development and management of IALS. These organizations were assisted by national research teams from the participating countries in developing the survey design. The survey design for the 1994 IALS is described below.

Target Population

The IALS target population was the civilian, noninstitutionalized population ages 16 to 65 in each country; however, countries were also permitted to sample older adults, and several did so. All IALS samples excluded full-time members of the military and people residing in institutions such as prisons, hospitals, and psychiatric facilities.

For the United States, the target population consisted specifically of civilian noninstitutionalized residents ages 16 to 65 in the 50 states and the District of Columbia, excluding members of the armed forces on active duty, those residing outside the United States, and those with no fixed household address (i.e., the homeless or residents of institutional group quarters, such as prisons and hospitals).

Sample Design

IALS was designed to provide data representative at the national level. Each country that participated in IALS agreed to draw a probability sample that would accurately represent its civilian, noninstitutionalized population ages 16 to 65. The final IALS sample design criteria specified that each country's sample should result in at least 1,000 respondents, the minimum sample size needed to produce reliable literacy proficiency estimates. Given the different sizes of the population of persons ages 16 to 65 in the countries involved, sample sizes varied considerably from country to country (ranging from 1,500 to 8,000 per country), but sample sizes were sufficiently large in all cases to support the estimation of reliable item parameters using Item Response Theory (IRT).

IALS countries were strongly encouraged to select high-quality probability samples because the use of probability designs would make it possible to produce unbiased estimates for individual countries and to compare these estimates across the countries. Because the available data sources and resources were different in each of the participating countries, however, no single sampling methodology was imposed. Each IALS country created its own sample design. All countries used probability sampling for at least some stages of their sample designs, and some used probability sampling for all stages of sampling. Sampling designs were approved by expert review.

The sample for the United States was selected from a sample of individuals in housing units who were completing their final round of interviews for the U.S. Census Bureau's Current Population Survey (CPS) in March, April, May, and June 1994. These housing units were included in the CPS for their initial interviews in December 1992 and January, February, and March 1993. The CPS is a large-scale continuous household survey of the civilian noninstitutionalized population age 15 and over. The frame for the CPS consisted of 1990 decennial census files, which are

continually updated for new residential construction and are adjusted for undercount, births, deaths, immigration, emigration, and changes in the armed forces.

The CPS sample is selected using a stratified multistage design. Housing units that existed at the time of the 1990 population census were sampled from the census list of addresses. Housing units that did not exist at that time were sampled from lists of new construction, when available, and otherwise by area sampling methods. Occupants of housing units that came into existence between the time of the CPS sample selection and the time of the IALS fieldwork had no chance of being selected for IALS.

The IALS sample was confined to 60 of the 729 CPS primary sampling units (PSUs). Within these 60 PSUs, all persons 16 to 65 years of age in the sampled housing units were classified into 20 cells defined by race/ethnicity and education. Within each cell, persons were selected for IALS with probability proportional to their CPS weights, with the aim of producing an equal probability sample of persons within cells. A total of 4,901 persons were selected for IALS. IALS interviews were conducted in October and November 1994.

Assessment Design

The success of IALS depended on the development and standardized application of a common set of survey instruments. The test framework explicitly followed the precedent set by the National Adult Literacy Survey, basing the test on U.S. definitions of literacy along three dimensions—prose literacy, document literacy, and quantitative literacy—but extending the instruments into an international context. Study managers from each participating country were encouraged to submit materials such as news articles and documents that could be used to create tasks with the goal of building a new pool of literacy tasks that could be linked to established scales. IALS field tested 175 tasks and identified 114 that were valid across cultures. Approximately half of these tasks were based on materials from outside North America. (However, each respondent was administered only a fraction of the pool of tasks, using a variant of matrix sampling.)

Each IALS country was given a set of model administration manuals and survey instruments as well as graphic files containing the pool of IALS literacy items with instructions to modify each item by translating the English text to its own language without altering the graphic representation. Certain rules governed the item modification process. For instance, some items required respondents to perform a task that was facilitated by the use of keywords. The keyword in the question might be identical to, similar but not exactly the same as, or a synonym of the word used in the body of the item, or respondents might be asked to choose among multiple keywords in the body of the item, only one of which was correct. Countries were required to preserve these conceptual associations during the translation process. Particular conventions used in the items—for example, currency units, date formats, and decimal delimiters—were adapted as appropriate for each country.

To ensure that the adaptation process did not compromise the psychometric integrity of the items, each country's test booklets were carefully reviewed for errors of adaptation. Countries were required to correct all errors found. However, this review was imperfect in two important respects. First, it is clear that countries chose not to incorporate a number of changes that were identified during the course of the review, believing that they "knew better." Second, the availability of empirical data from the study has permitted the identification of several additional sources of task and item difficulty that were not included in the original framework, which was based on research by Irwin Kirsch of ETS and Peter Mosenthal of Syracuse University. (See 1990 publication, "Exploring Document Literacy: Variables Underlying the Performance of Young Adults," by I.S. Kirsch and P.B. Mosenthal, in *Reading Research Quarterly* 25: 5–30.) Item adaptation guidelines and item review procedures associated with subsequent rounds of IALS data collection were adapted to reflect this additional information.

The model background questionnaires contained two sets of questions: mandatory questions, which all countries were required to include; and optional questions, which were recommended but not required. Countries were not required to field literal translations of the mandatory questions, but were asked to respect the conceptual intent of each question in adapting it for use. Countries were permitted to add questions to their background questionnaires if the additional burden on respondents would not reduce response rates. Statistics Canada reviewed all background questionnaires (except Sweden's) before the pilot survey and offered comments and suggestions to each country.

Data Collection and Processing

IALS data for the first round of countries were collected through in-person household interviews in the fall of 1994. Each country mapped its national dataset into a highly structured, standardized record layout that it sent to Statistics Canada. Further description follows.

Reference dates. Respondents answered questions about jobs they may have held in the 12 months before the survey was administered.

Data collection. Statistics Canada and ETS coordinated the development and management of IALS. Participating countries were given model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments and for handling nonresponse codings.

Countries were permitted to adapt these models to their own national data collection systems, but they were required to retain a number of key features: (1) respondents were to complete the core and main test booklets alone, in their homes, without help from another person or from a calculator; (2) respondents were not to be given monetary incentives for participating; (3) despite the prohibition on monetary incentives, interviewers were provided with procedures to maximize the number of completed background questionnaires and were to use a common set of coding specifications to deal with nonresponse. This last requirement was critical. Because noncompletion of the core and main task booklets was correlated with ability, background information about nonrespondents was needed in order to impute cognitive data for these persons.

IALS countries were instructed to obtain at least a background questionnaire from sampled individuals. All countries participating in IALS instructed interviewers to make callbacks at households that were difficult to contact.

In general, the survey was carried out in the national language. In Canada, respondents were given a choice of English or French, and in Switzerland, samples drawn from French-speaking and German-speaking cantons were required to respond in those respective languages. When respondents could not speak the designated language, attempts were made to complete the background questionnaire so that their literacy level could be estimated and the possibility of distorted results would be reduced. In the United States, the test was given in English, but a Spanish version of the background questionnaire and bilingual interviewers were available to assist individuals whose native language was not English.

Survey respondents spent approximately 20 minutes answering a common set of background questions concerning their demographic characteristics, educational experiences, labor market experiences, and literacy-related activities. Responses to these background questions made it possible to summarize the survey results using an array of descriptive variables, and also increased the accuracy of the proficiency estimates for various subpopulations. After answering the background questions, the remainder of respondents’ time was spent completing a booklet of literacy tasks designed to measure their prose, document, and quantitative skills. Most of these tasks were open-ended, requiring respondents to provide a written answer.

In the United States, the IALS interview period was from October to November 1994. IALS was conducted by 149 Census Bureau interviewers. All of them had at least 5 days of interviewer training. They were given a one-day training on IALS and were provided with substantial training and reference materials based on the Canadian training package. They also performed a day of field training under the supervision of a regional office supervisor. Each interviewer had an average workload of 33 interviews, and the average number of response interviews per interviewer was 21. They were supervised by six regional supervisors who reviewed and commented on their work.

Before data collection, a letter was sent to the selected addresses describing the upcoming survey. The survey was limited to 90 minutes. If a respondent took more than 20 minutes per block, the interviewer was instructed to move the respondent on to the next block.

Data processing. As a condition of their participation in IALS, countries were required to capture and process their files using procedures that ensured logical consistency and acceptable levels of data capture error. Specifically, countries were advised to conduct complete verification of the captured scores (i.e., enter each record twice) in order to minimize error rates. One hundred percent keystroke validation was needed. Specific details about scoring are provided in a separate section below.

To create a workable comparative analysis, each IALS country was required to map its national dataset into a highly structured, standardized record layout. In addition to specifying the position, format, and length of each field, this International Record Layout included a description of each variable and indicated the categories and codes to be provided for that variable. Upon receiving a country’s file, Statistics Canada performed a series of range checks to ensure compliance to the prescribed format. When anomalies were detected, countries corrected the problems and submitted new files. Statistics Canada did not, however, perform any logic or flow edits, as it was assumed that participating countries performed this step themselves.

Editing. Most countries followed IALS guidelines, verifying 100 percent of their data capture operation. The two countries that did not comply with this recommendation conducted sample verifications, one country at 20 percent and the other at 10 percent. Each country coded and edited its own data, mapping its national dataset into the detailed International Record Layout, which included a description of each variable and indicated the categories and codes to be provided for that variable. Industry, occupation, and education were coded using the standard international coding schemes: the International Standard Industrial Classification (ISIC), the International Standard Classification of Occupations (ISCO), and the International Standard Classification of Education (ISCED). Coding schemes were provided for open-ended items; the coding schemes came with specific instructions so that coding error could be contained to acceptable levels.

Scoring. Respondents' literacy proficiencies were estimated based on their performance on the cognitive tasks administered in the assessment. Because the open-ended items used in IALS elicited a large variety of responses, responses had to be grouped in order to summarize the performance results. As they were scored, responses to IALS open-ended items were classified as correct, incorrect, or omitted. The models employed to estimate ability and difficulty were predicated on the assumption that the scoring rubrics developed for the assessment were applied in a consistent fashion within and between countries. To reinforce the importance of consistent scoring, a meeting of national study managers and chief scorers was held prior to the commencement of scoring for the main study. The group spent 2 days reviewing the scoring rubrics for all the survey items. Where this review uncovered ambiguities and situations not covered by the guides, clarifications were agreed to collectively, and these clarifications were then incorporated into the final rubrics. To provide ongoing support during the scoring process, Statistics Canada and ETS maintained a joint scoring hotline. Any scoring problems encountered by chief scorers were resolved by this group, and decisions were forwarded to all national study managers. Study managers conducted intensive scoring training using the scoring manual and discussed unusual responses with scorers. They also offered additional training to some scorers, as needed, to raise their accuracy to the level achieved by other scorers.

To maintain coding quality within acceptable levels of error, each country undertook to rescore a minimum of 10 percent of all assessments. Where significant problems were encountered, larger samples of a particular scorer's work were to be reviewed and, where necessary, their entire assignments rescored. Countries were not required to resolve contradictory scores in the main survey (as they had been in the pilot), since outgoing agreement rates were far above minimum acceptable tolerances.

Since there could still be significant differences in the consistency of scoring between countries, countries agreed to exchange at least 300 randomly selected booklets with another country sharing the same test language. In all cases where serious discrepancies were identified, countries were required to rescore entire items or discrepant code pairs.

Intra-country rescoring. A variable sampling ratio procedure was set up to monitor scoring accuracy. At the beginning of scoring, almost all responses were rescored to identify inaccurate scorers and to detect unique or difficult responses that were not covered in the scoring manual. After a satisfactory level of accuracy was achieved, the rescoring ratio was dropped to a maintenance level to monitor the accuracy of all scorers. Average agreements were calculated across all items. Precautions were taken to ensure that the first and second scores were truly independent.

Intercountry rescoring. To determine intercountry scoring reliabilities for each item, the responses of a subset of examinees were scored by two separate groups. Usually, these scoring groups were from different countries. Intercountry score reliabilities were calculated by Statistics Canada, and then evaluated by ETS. Based on the evaluation, every country was required to introduce a few minor changes in scoring procedures. In some cases, ambiguous instructions in the scoring manual were found to be causing erroneous interpretations and therefore lower reliabilities.

Using the intercountry score reliabilities, researchers could identify poorly constructed items, ambiguous scoring criteria, erroneous translations of items or scoring criteria, erroneous printing of items or scoring criteria, scorer inaccuracies, and, most important, situations in which one country consistently scored differently from another. In the latter circumstance, scorers in one country may consistently rate a certain response as being correct while those in another country score the same response as incorrect. ETS and Statistics Canada examined scoring carefully to identify situations in which scorers in one country were consistently rating a certain response as being correct while those in another country were scoring the same response as incorrect. Where a systematic error was identified in a particular country, the original scores for that item were corrected for the entire sample.

Estimation Methods

Weighting was used in the 1994 IALS to adjust for sampling and nonresponse. Responses to the literacy tasks were scored using IRT scaling. A multiple imputation procedure based on plausible values methodology was used to estimate the literacy proficiencies of individuals who completed literacy tasks.

Weighting. IALS countries used different methods for weighting their samples. Countries with known probabilities of selection could calculate a base weight using the probability of selection. To adjust for unit nonresponse, all countries poststratified their data to known population counts, and a comparison of the distribution of the age and sex characteristics of the actual and weighted samples indicates that the samples were comparable to the overall populations of IALS countries. Another commonly used approach was to weight survey data to adjust the rough estimates produced by the sample to match known population counts from sources external to IALS. This "benchmarking" procedure assumes that the characteristics of nonrespondents are similar to those of respondents. It is most effective when the variables used for benchmarking are strongly correlated with the characteristic of interest—in this case, literacy levels. For IALS, the key benchmarking variables were age, employment status, and education. All of the IALS countries benchmarked to at least one of these variables. The United States used education.

Weights for the U.S. IALS sample included two components. The first assigned weights to CPS respondents, and the second assigned weights to IALS respondents.

The CPS weighting scheme was a complex one involving three components: basic weighting, noninterview adjustment, and ratio adjustment. The basic weighting compensated for unequal selection probabilities. The noninterview adjustment compensated for nonresponse within weighting cells created by clusters of PSUs of similar size; Metropolitan Statistical Area (MSA) clusters were subdivided into central city areas, and the balance of the MSA and non-MSA clusters were divided into urban and rural areas. The ratio adjustment made the weighted sample distributions conform to known distributions on such characteristics as age, race, Hispanic origin, sex, and residence.

The weights of persons sampled for IALS were adjusted to compensate for the use of the four rotation groups, the sampling of the 60 PSUs, and the sampling of persons within the 60 PSUs. The IALS noninterview adjustment compensated for sampled persons for whom no information was obtained because they were absent, refused to participate, had a short-term illness, had moved, or had experienced an unusual circumstance that prevented them from being interviewed. Finally, the IALS ratio adjustment ensured that the weighted sample distributions across a number of education groups conformed to March 1994 CPS estimates of these numbers.

Scaling. The scaling model used in IALS was the two-parameter logistic model based on IRT.

Items developed for IALS were based on the framework used in three previous large-scale assessments: the Young Adult Literacy Assessment, the DOL survey, and the National Adult Literacy Survey. As a result, IALS items shared the same characteristics as the items in these earlier surveys. The English versions of IALS items were reviewed and tested to determine whether they fit into the literacy scales in accordance with the theory and whether they were consistent with the National Adult Literacy Survey data. Quality control procedures for item translation, scoring, and scaling followed the same procedures used in the National Adult Literacy Survey and extended the methods used in other international studies.

Identical item calibration procedures were carried out separately for each of the three literacy scales: prose, document, and quantitative literacy. Using a modified version of Mislevy and Bock's 1982 BILOG computer program—see *BILOG: Item analysis and test scoring with binary logistic models*, Scientific Software—the two-parameter logistic IRT model was fit to each item using sample weights. BILOG procedures are based on an extension of the marginal-maximum-likelihood approach described by Bock and Aitkin in their 1981 *Psychometrika* article, "Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm."

Most of the items administered in IALS were successful from a psychometric standpoint. However, despite stringent efforts at quality control, some of the assessment items did not meet the criteria for inclusion in the final tabulation of results. Specifically, in carrying out the IRT modeling used to create the three literacy scales, researchers found that a number of assessment items had significantly different item parameters across IALS countries.

Imputation. A respondent had to complete the back-ground questionnaire, pass the core block of literacy tasks, and attempt at least five tasks per literacy scale

in order for researchers to be able to estimate his or her literacy skills directly. Literacy proficiency data were imputed for individuals who failed or refused to perform the core literacy tasks and for those who passed the core block but did not attempt at least five tasks per literacy scale. Because the model used to impute literacy estimates for nonrespondents relied on a full set of responses to the background questions, IALS countries were instructed to obtain at least a background questionnaire from sampled individuals. IALS countries were also given a detailed nonresponse classification to use in the survey.

Literacy proficiencies of respondents were estimated using a multiple imputation procedure based on plausible values methodology. Special procedures were used to impute missing cognitive data.

Literary proficiency estimation (plausible values). A multiple imputation procedure based on plausible values methodology was used to estimate respondents' literacy proficiency in the 1994 IALS. When a sampled individual decided to stop the assessment, the interviewer used a standardized nonresponse coding procedure to record the reason why the person was stopping. This information was used to classify nonrespondents into two groups: (1) those who stopped the assessment for literacy-related reasons (e.g., language difficulty, mental disability, or reading difficulty not related to a physical disability); and (2) those who stopped for reasons unrelated to literacy (e.g., physical disability or refusal). About 45 percent of the individuals did not complete the assessment for reasons related to their literacy skills; the other respondents gave no reason for stopping or gave reasons unrelated to their literacy.

When individuals cited a literacy-related reason for not completing the cognitive items, it implies that they were unable to respond to the items. On the other hand, citing reasons unrelated to literacy implies nothing about a person's literacy proficiency. Based on these interpretations, IALS adapted a procedure originally developed for the National Adult Literacy Survey to treat cases in which an individual responded to fewer than five items per literacy scale, as follows: (1) if the individual cited a literacy-related reason for not completing the assessment, then all consecutively missing responses at the end of the block of items were treated as wrong; and (2) if the individual cited reasons unrelated to literacy for not completing the assessment, then all consecutively missing responses at the end of a block were treated as "not reached."

Proficiency values were estimated based on respondents' answers to the background questions and the cognitive items. As an intermediate step, the functional relationship between these two sets of information was calculated, and this function was used to obtain unbiased proficiency estimates with reduced error variance. A respondent's proficiency was calculated from a posterior distribution that was the multiple of two functions: a conditional distribution of proficiency, given responses to the background questions; and a likelihood function of proficiency, given responses to the cognitive items.

Recent Changes

Since IALS was a one-time assessment, there are no changes to report.

Future Plans

There are no plans to conduct IALS again. However, a new survey, the Adult Literacy and Lifeskills Survey (ALL), was administered in 2003 (see chapter 24). The aspects of this survey that address literacy were built on methodologies used in IALS.

5. DATA QUALITY AND COMPARABILITY

The literacy tasks contained in IALS and the adults asked to participate in the survey were samples drawn from their respective universes. As such, they were subject to some measurable degree of uncertainty. IALS implemented procedures to minimize both sampling and nonsampling errors. The IALS sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. Scientific procedures employed in the study design and the scaling of literacy tasks permitted a high degree of confidence in the resulting estimates of task difficulty. Quality control activities continued during interviewer training, data collection, and processing of the survey data.

In addition, special evaluation studies were conducted to examine issues related to the quality of IALS. These studies included (1) an external evaluation of IALS methodology; (2) an examination of how similar or different the sampled persons were from the overall population; (3) an evaluation of the extent to which the literacy levels of the population in the database for each nation were predictable based on demographic characteristics; (4) an examination of the assumption of unidimensionality; and (5) an evaluation of the construct validity of the adult literacy scales.

Sampling Error

Because IALS employed probability sampling, the results were subject to sampling error. Although small, this error was higher in IALS than in most studies because the cost of surveying adults in their homes is so high. Most countries simply could not afford large sample sizes.

Each country provided a set of replicate weights for use in a jackknife variance estimation procedure.

There were three situations in which nonprobability-based sampling methods were used: France and Germany used “random route” procedures for selecting households into their samples, and Switzerland used an alphabetic sort to select one member of each household. However, based on the available evidence, it is not believed that these practices introduced significant bias into the survey estimates.

In 1998, the U.K. Office of National Statistics coordinated the European Adult Literacy Review, a split-sample survey intended, in part, to measure the effects of sampling methods on the IALS results. This follow-up survey compared an IALS sample design with an alternative, standardized “best practice” design. Although certain differences were noted between the two samples, the IALS sample design was not confirmed to be inferior to the “best practice” design.

Nonsampling Error

The key sources of nonsampling error in the 1994 IALS were differential coverage across countries and nonresponse bias, which occurred when different groups of sampled individuals failed to participate in the survey. Other potential sources of nonsampling error included deviations from prescribed data collection procedures and errors of logic that resulted from mapping idiosyncratic national data into a rigid international format. Scoring error, associated with scoring open-ended tasks reliably within and between countries, also occurred. Finally, because IALS data were collected and processed independently by the various countries, the study was subject to uneven levels of commonplace data capture, data processing, and coding errors.

Three studies were conducted to examine the possibility of nonresponse bias. Because the sampling frames for Canada and the United States contained information about the characteristics of sampled individuals, it was possible to compare the characteristics of respondents and nonrespondents, particularly with respect to literacy skill profiles. The Swedish National Study Team also commissioned a nonresponse follow-up study.

Coverage error. The design specifications for IALS stated that in each country the study should cover the civilian, noninstitutionalized population ages 16 to 65. It is the usual practice to exclude the institutional population from national surveys because of the difficulties in conducting interviews in institutional settings. Similarly, it is not uncommon to exclude certain other parts of a country’s population that pose difficult survey problems (e.g., persons living in sparsely populated areas). The intended coverage of the surveys generally conformed well to the design specifications: each of the IALS countries attained a high level of population coverage, ranging from a low of 89 percent in Switzerland to a high of 99 percent in the Netherlands and Poland. However, it should be noted that actual coverage is generally lower than the intended coverage because of deficiencies in sampling frames and sampling frame construction (e.g., failures to list some households and some adults within listed households). In the United States, for example, comparing population sizes estimated from the survey with external benchmark figures suggests that the overall coverage rate for the CPS (the survey from which the IALS sample was selected) is about 93 percent, but that it is much lower for certain population subgroups (particularly young Black male adults).

Nonresponse error. For IALS, several procedures were developed to reduce biases due to nonresponse, based on how much of the survey the respondent completed.

Unit nonresponse. The definition of a respondent for IALS was a person who partially or fully completed the background questionnaire. Unweighted response rates varied considerably from country to country, ranging from a high of 69 percent (Canada, Germany) to a low of 45 percent (the Netherlands), with four countries in the 55–60 percent range.

In the United States, which had a response rate of 60 percent, nonresponse to IALS occurred for two reasons: (1) some individuals did not respond to the CPS; and (2) some of the CPS respondents selected for IALS did not respond to the IALS instruments. In any given month, nonresponse to the CPS is typically quite low, around 4 to 5 percent. Its magnitude in the expiring rotation groups employed for IALS selection is not known. About half of the CPS nonresponse is caused by refusals to participate, while the remainder is caused by temporary absences, other failures to contact individuals, the inability of individuals contacted to respond, and unavailability for other reasons.

A sizable proportion of the nonresponse to the IALS background questionnaire was attributable to persons who had moved. For budgetary reasons, it was decided

that persons who were not living at the CPS addresses at the time of the IALS interviews would not be contacted. This decision had a notable effect on the sample of students, who are sampled in dormitories and other housing units in the CPS only if they do not officially reside at their parents' homes. Those who reside at their parents' homes are included in the CPS at that address, but because most of these students were away at college during the IALS interview period (October to November 1994), they could not respond to IALS.

The high level of nonresponse for college students could cause a downward bias in the literacy skill-level estimates. This group represents only a small proportion of the U.S. population, however, so the potential bias is likely to be quite small. Furthermore, a comparison of IALS results to the U.S. National Adult Literacy Survey data discounts this as a major source of bias.

Item nonresponse. The weighted percentage of omitted responses for the U.S. IALS sample ranged from 0 to 18 percent.

Not-reached responses were classified into two groups: nonparticipation immediately or shortly after the background information was collected; and premature withdrawal from the assessment after a few cognitive items were attempted. The first type of not-reached response varied a great deal across countries according to the frames from which the samples were selected. The second type of not-reached response was due to quitting the assessment early, resulting in incomplete cognitive data. Not-reached items were treated as if they provided no information about the respondent's proficiency, so they were not included in the calculation of likelihood functions for individual respondents. Therefore, not-reached responses had no direct impact on the proficiency estimation for subpopulations. The impact of not-reached responses on the proficiency distributions was mediated through the subpopulation weights.

Measurement error. Assessment tasks were selected to ensure that, among population subgroups, each literacy domain (prose, document, and quantitative) was well covered in terms of difficulty, stimuli type, and content domain. The IALS item pool was developed collectively by participating countries. Items were subjected to a detailed expert analysis at ETS and vetted by participating countries to ensure that the items were culturally appropriate and broadly representative of the population being tested. For each country, experts who were fluent in both English and the language of the test reviewed the items and identified ones that had been improperly adapted. Countries were asked to correct problems detected during this review process. To ensure that all of the final survey items had a high probability of functioning well, and to familiarize participants with the unusual operational requirements involved in data collection, each country was required to conduct a pilot survey. Although the pilot surveys were small and typically were not based strictly on probability samples, the information they generated enabled ETS to reject items, to suggest modifications to a few items, and to choose good items for the final assessment. ETS's analysis of the pilot survey data and recommendations for the final test design were presented to and approved by participating countries.

Data Comparability

While most countries closely followed the data collection guidelines provided, some did deviate from the instructions. First, two countries (Sweden and Germany) offered participation incentives to individuals sampled for their survey. The incentive paid was trivial, however, and it is unlikely that this practice distorted the data. Second, the doorstep introduction provided to respondents differed somewhat from country to country. Three countries (Germany, Switzerland, and Poland) presented the literacy test booklets as a review of the quality of published documents rather than as an assessment of the respondent's literacy skills. A review of these practices suggested that they were intended to reduce response bias and were warranted by cultural differences in respondents' attitudes toward being tested. Third, there were differences across the countries in the way in which interviewers were paid. No guidelines were provided on this subject, and the study teams therefore decided what would work best in their respective countries. Fourth, several countries adopted field procedures that undermined the objective of obtaining completed background questionnaires for an overwhelming majority of selected respondents.

This project was designed to produce data comparable across cultures and languages. After one of the countries in the first round raised concerns about the international comparability of the survey data, Statistics Canada decided that the IALS methodology should be subjected to an external evaluation. In the judgment of the expert reviewers, the considerable efforts that were made to develop standardized survey instruments for the different nations and languages were successful, and the data obtained from them should be broadly comparable.

However, the standardization of procedures with regard to other aspects of survey methodology was not

achieved to the extent desired, resulting in several weaknesses. Nonresponse proved to be a particular weakness, with generally very high nonresponse rates and variation in nonresponse adjustment procedures across countries. For some countries the sample design was problematic, resulting in some unknown biases. The data collection and its supervision differed between participating countries, and some clear weaknesses were evident for some countries. The reviewers felt that the variation in survey execution across countries was so large that they recommended against publication of comparisons of overall national literacy levels. They did, however, despite the methodological weaknesses, recommend that the survey results be published. They felt that the instruments developed for measuring adult literacy constituted an important advance, and the results obtained for the instruments in the first round of IALS were a valuable contribution to the field. They recommended that the survey report focus on analyses of the correlates of literacy (e.g., education, occupation, and age) and the comparison of these correlates across countries. Although these analyses might also be distorted by methodological problems, they believed that the analyses were likely to be less affected by these problems than were the overall literacy levels.

6. CONTACT INFORMATION

For content information on IALS, contact:

Eugene Owen
Phone: (202) 502-7422
E-mail: eugene.owen@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

Murray, T.S., Kirsch, I.S. and Jenkins, L.B. (eds.). (1997). *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey* (NCES 98-053). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Chapter 24: Adult Literacy and Lifeskills Survey (ALL)

1. OVERVIEW

The Adult Literacy and Lifeskills Survey (ALL) is an international comparative study designed to provide participating countries, including the United States, with information about the skills of their adult populations ages 16 to 65. The development and management of the study were coordinated by Statistics Canada and the Educational Testing Service (ETS) in collaboration with the National Center for Education Statistics (NCES) of the U.S. Department of Education; the Organization for Economic Cooperation and Development (OECD); the Regional Office of Education for Latin America and the Caribbean (OREALC); and the Institute for Statistics (UIS) of the United Nations Educational, Scientific, and Cultural Organization (UNESCO).

ALL measured the literacy and numeracy skills of a nationally representative sample from each participating country. On a pilot basis, ALL also measured adults' problem-solving skills and gathered information on their familiarity with information and communication technology (ICT). ALL builds on the foundation of earlier studies of adult literacy. Chief among these earlier studies is the International Adult Literacy Survey (IALS), which was conducted in three phases (1994, 1996, and 1998) in 20 nations, including the United States. The following six countries participated in ALL: Italy, Norway, Switzerland, Bermuda, Canada, and the United States.

ADULT LITERACY AND LIFESKILLS SURVEY

ALL collected:

- Background assessments
- Literacy assessments in prose literacy, document literacy, numeracy, and problem-solving domains

Purpose

To (1) profile and compare the literacy skills in adult populations; (2) profile and compare the level and distribution of directly assessed numeracy skills among adult populations in participating countries; (3) profile and compare the level and distribution of problem-solving skills among the adult populations of the countries surveyed; and (4) collect comparable data on participation in formal adult education.

Components

Each ALL country was given a set of model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments. ALL instruments consisted of three parts: (1) a background questionnaire, which collected demographic information about respondents; (2) a set of core literacy tasks, which screened out respondents with very limited literacy skills; and (3) a main booklet of literacy tasks, used to calibrate literacy levels.

Background Questionnaire. The background questionnaire collected general participant information (such as sex, age, race/ethnicity, education level, and labor force status) and posed more targeted questions related to literacy practices, familiarity with ICT, education coursetaking, and health.

Literacy Assessment.
Core literacy tasks. The core literacy tasks were presented to respondents once they had completed the background questionnaire. The booklet for the core literacy tasks contained six simple tasks. Only those who answered at least two of the core tasks correctly were given the full assessment.

Main literacy tasks. The main literacy tasks for the ALL psychometric assessment consisted of tasks in prose literacy, document literacy, numeracy, and problem-solving domains. The assessment included four 30-minute blocks of literacy items (i.e., prose and document literacy), two 30-minutes blocks of numeracy items, and two 30-minute blocks of problem-solving items. A four-domain ALL assessment was implemented in Bermuda, Canada, Italy, Norway, and the French- and German-language regions of Switzerland. The United States and the Italian-language region of Switzerland carried out a three-domain ALL assessment that excluded the problem-solving domain. The blocks of assessment items were organized into 28 task booklets in the case of the four-domain assessment and into 18 task booklets for the three-domain assessment. The assessment blocks were distributed to the task booklets according to a balanced incomplete block (BIB) design whereby each task booklet contained two blocks of items.

Periodicity
ALL was conducted between the fall of 2003 and early spring 2004. In the United States, data collection for the main study took place between January and June 2003.

2. USES OF DATA

ALL sought to provide researchers with information on skill gain and loss in the adult population. This was achieved through the measurement of prose and document literacy. Furthermore, the study extended the range of skills measured by adding tasks for problem-solving, numeracy, and ICT skills. This allows researchers to examine the profiles of important foundation skills. The study makes it possible to explore the interrelationships among skill domains as well as their links to major antecedents and outcomes, such as the quantity and quality of initial education and the impact of skills on employability, wages, and health.

In addition, information from ALL addresses questions such as the following:

- What is the distribution of literacy and numeracy skills among American adults? How do these skill distributions compare to those of other countries?

- What is the relationship between these literacy skills and the economic, social, and personal characteristics of individuals? For example: Do different age or linguistic groups manifest different skill levels? Do males and females perform differently? At what kinds of jobs do people at various literacy levels work? What wages do they earn? How do adults who have completed different levels of education perform?

- What is the relationship between these skills and the economic and social characteristics of nations? For example, how do the skills of the adult labor force of a country match up with areas of the economy that are growing?

3. KEY CONCEPTS

Four skill domains are conceptualized in ALL: prose literacy, document literacy, numeracy, and problem solving. Two of them, namely, prose and document literacy, are defined and measured in the same manner as in IALS (see chapter 23). Numeracy and problem solving are new domains.

Prose literacy. The knowledge and skills needed to understand and use information from texts, including editorials, news stories, brochures, and instruction manuals.

Document literacy. The knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables, and charts.

Numeracy. The knowledge and skills required to effectively manage the mathematical demands of diverse situations.

Problem solving. Problem solving involves goal-directed thinking and action in situations for which no routine solution procedure is available. The problem solver has a more or less well-defined goal, but does not immediately know how to reach it. The incongruence of goals and admissible operators constitutes a problem. The understanding of the problem situation and its step-by-step transformation

based on planning and reasoning constitute the process of problem solving.

Literacy scale. For each skill assessment domain, proficiency is denoted on a scale ranging from 0 to 500 points. Each score denotes a point at which a person has an 80 percent chance of successfully completing tasks that are associated with a similar level of difficulty. For the prose and document literacy domains as well as the numeracy domain, experts defined five broad levels of difficulty, each corresponding to a range of scores. For the problem-solving domain, experts defined four broad levels of difficulty.

4. SURVEY DESIGN

Each participating country was required to design and implement the Adult Literacy and Lifeskills Survey according to specified guidelines and standards. These ALL standards established the minimum survey design and implementation requirements for the following project areas: survey planning; target population; method of data collection; sample frame; sample design; sample selection; literacy assessment design; background questionnaire; task booklets; instrument requirements to facilitate data processing; data collection; respondent contact strategy; response rate strategy; interviewer hiring, training, and supervision; data capture; coding and scoring; data file format and editing; weighting; estimation; confidentiality; survey documentation; and pilot survey.

Target Population

Each participating country designed a sample to be representative of its civilian noninstitutionalized population ages 16 to 65 (inclusive). Countries were also at liberty to include adults over the age of 65 in the sample provided that a minimum suggested sample size requirement was satisfied for the 16 to 65 age group. Canada opted to include in its target population adults over the age of 65. All of the remaining countries restricted the target population to the 16 to 65 age group. Exclusions from the target population for practical operational reasons were acceptable provided a country's survey population did not differ from the target population by more than 5 percent (i.e., provided that the total number of exclusions from the target population due to undercoverage was not more than 5 percent of the target population). All countries indicate that this 5 percent requirement was satisfied. Each country chose or developed a sample frame to cover the target population.

Sample Design

Each participating country was required to use a probability sample representative of the national population ages 16 to 65. A sample size of 5,400 completed cases in each official language was recommended for each country that was implementing the full ALL psychometric assessment (i.e., comprising the prose literacy, document literacy, numeracy, and problem-solving domains). A sample size of 3,420 complete cases in each official language was recommended if the problem-solving domain was excluded from the ALL assessment.

The available sampling frames and resources varied from one country to another. Therefore, the particular probability sample design to be used was left to the discretion of each country. Each country's proposed sample design was reviewed by Statistics Canada to ensure that the sample design standards and guidelines were satisfied.

A stratified multistage probability sample design was employed in the United States. The first stage of sampling consisted of selecting a sample of 60 primary sampling units (PSUs) from a total of 1,880 PSUs that were formed using a single county or a group of contiguous counties, depending on the population size and the area covered by a county or counties. The PSUs were stratified on the basis of the social and economic characteristics of the population, as reported in the 2000 census. The following characteristics were used to stratify the PSUs: region of the country, Metropolitan Statistical Area (MSA), population size, percentage of African-American residents, percentage of Hispanic residents, and per capita income. The largest PSUs in terms of population size were included in the sample with certainty. For the remaining PSUs, one PSU per stratum was selected with probability proportional to the population size.

At the second sampling stage, a total of 505 geographic segments were systematically selected with probability proportional to population size from the sampled PSUs. Segments consist of area blocks (as defined by the 2000 census) or combinations of two or more nearby blocks. They were formed to satisfy criteria based on population size and geographic proximity. The third stage of sampling involved the listing of the dwellings in the selected segments and the subsequent selection of a random sample of dwellings. An equal number of dwellings was selected from each sampled segment. At the fourth and final stage of sampling, one eligible person was randomly selected within households with fewer than four eligible adults. In households with four or more eligible persons, two adults were randomly selected.

Assessment Design

A BIB assessment design was used to measure the skill domains. The BIB design comprised a set of assessment tasks organized into smaller sets of tasks, or blocks. Each block contained assessment items from one of the skill domains and covered a wide range of difficulty (i.e., from easy to difficult). The blocks of items were organized into task booklets according to a BIB design. Individual respondents were not required to take the entire set of tasks. Instead, each respondent was randomly administered one of the task booklets.

ALL assessment. The ALL psychometric assessment consisted of the prose literacy, document literacy, numeracy, and problem-solving domains. The assessment included four 30-minute blocks of literacy items (i.e., prose and document literacy), two 30-minute blocks of numeracy items, and two 30-minute blocks of problem-solving items. A four-domain ALL assessment was implemented in Bermuda, Canada, Italy, Norway, and the French- and German-language regions of Switzerland. The United States and the Italian-language region of Switzerland carried out a three-domain ALL assessment that excluded the problem-solving domain.

The blocks of assessment items were organized into 28 task booklets in the four-domain assessment and into 18 task booklets in the three-domain assessment. The assessment blocks were distributed to the task booklets according to a BIB design whereby each task booklet contained two blocks of items. The task booklets were randomly distributed among the selected sample. In addition, the data collection activity was closely monitored in order to obtain approximately the same number of complete cases for each task booklet, except for two-task booklets in the three-domain assessment containing only numeracy items, which required a larger number of complete cases.

Data Collection and Processing

The data collection for the ALL project took place between the fall of 2003 and early spring 2004, depending on the country. However, in the United States, data collection for the main study took place between January and June 2003. In the United States, a nationally representative sample of 3,420 adults ages 16 to 65 participated in ALL. Trained interviewers administered approximately 45 minutes of background questions and 60 minutes of assessment items to participants in their homes.

Reference dates. Respondents answered questions about jobs they may have held in the 12 months before the survey was administered.

Data collection. The ALL survey design combined educational testing techniques with those of household survey research to measure literacy and provide the information necessary to make these measures meaningful. The respondents were first asked a series of questions to obtain background and demographic information on educational attainment, literacy practices at home and at work, labor force information, ICT use, adult education participation, and literacy self-assessment. Once the background questionnaire had been completed, the interviewer presented a booklet containing six simple tasks (the core tasks). Respondents who passed the core tasks were given a much larger variety of tasks, drawn from a pool of items grouped into blocks; each booklet contained two blocks that represented about 45 items. No time limit was imposed on respondents, and they were urged to try each item in their booklet. Respondents were given the maximum leeway to demonstrate their skill levels, even if their measured skills were minimal.

To ensure high-quality data, ALL guidelines specified that each country should work with a reputable data collection agency or firm, preferably one with its own professional, experienced interviewers. The interviews were to be conducted in the home in a neutral, nonpressured manner. Interviewer training and supervision was to be provided that emphasized the selection of one person per household (if applicable), the selection of one of the 28 main task booklets (if applicable), the scoring of the core task booklet, and the assignment of status codes. Finally, the interviewers' work was to be supervised by the use of quality checks—frequent quality checks at the beginning of the data collection and fewer quality checks throughout the remainder of the data collection—and by having help available to interviewers during entire the data collection period.

Several precautions were taken against nonresponse bias. Interviewers were specifically instructed to return several times to nonrespondent households in order to obtain as many responses as possible. In addition, all countries were asked to ensure that the address information provided to interviewers was as complete as possible in order to reduce potential household identification problems. Countries were asked to complete a debriefing questionnaire after the study in order to demonstrate that the guidelines had been followed, as well as to identify any collection problems they had encountered.

The United States administered the survey only in English. It used 106 interviewers during the data collection process, assigning approximately 64 cases to each interviewer. Professional interviewers were used

to conduct the survey, although approximately one-quarter of the interviewers had no previous survey experience.

Data processing. As a condition of their participation in ALL, countries were required to capture and process their files using procedures that ensured logical consistency and acceptable levels of data capture error. Specifically, countries were advised to conduct complete verification of the captured scores (i.e., enter each record twice) in order to minimize error rates. Because the process of accurately capturing the task scores is essential to high data quality, 100 percent keystroke verification was required.

Each country was also responsible for coding industry, occupation, and education using standard coding schemes, such as the International Standard Industrial Classification (ISIC), the International Standard Classification of Occupations (ISCO), and the International Standard Classification of Education (ISCED). Coding schemes were provided by Statistics Canada for all open-ended items, and countries were given specific instructions about the coding of such items.

In order to facilitate comparability in data analysis, each ALL country was required to map its national dataset into a highly structured, standardized record layout. In addition to specifying the position, format, and length of each field, the international record layout included a description of each variable and indicated the categories and codes to be provided for that variable. Upon receiving a country's file, Statistics Canada performed a series of range checks to ensure compliance to the prescribed format; flow and consistency edits were also run on the file. When anomalies were detected, countries were notified of the problem and were asked to submit cleaned files.

Scoring. Persons in each country charged with scoring received intense training, using the ALL scoring manual, in scoring responses to the open-ended items. They were also provided a tool for capturing closed format questions. To aid in maintaining scoring accuracy and comparability between countries, ALL introduced the use of an electronic bulletin board where countries could post their scoring questions and receive scoring decisions from the domain experts. This information could be seen by all countries, who could then adjust their scoring.
To further ensure quality, countries were monitored as to the quality of their scoring in two ways.

First, within a country, at least 20 percent of the tasks had to be rescored. Guidelines for intra-country rescoring involved rescoring a larger portion of booklets at the beginning of the scoring process to identify and rectify as many scoring problems as possible. In a second phase, countries selected a smaller portion of the next third of the scoring booklets; this phase was viewed as a quality monitoring measure and involved rescoring a smaller portion of booklets regularly to the end of the rescoring activities. The two sets of scores needed to match with at least 95 percent accuracy before the next step of processing could begin. In fact, most of the intra-country scoring reliabilities were above 95 percent. Where errors occurred, a country was required to go back to the booklets and rescore all the questions with problems and all the tasks that belonged to a problem scorer.

Second, an international rescore was performed. Each country had 10 percent of its sample rescored by scorers in another country. For example, a sample of task booklets from the United States was rescored by the persons who had scored Canadian English booklets, and vice versa. The main goal of the rescore was to verify that no country scored consistently differently from another country. Intercountry score reliabilities were calculated by Statistics Canada and the results were evaluated by the ETS. Again, strict accuracy was demanded: a 90 percent correspondence was required before the scores were deemed acceptable. Any problems detected had to be rescored.

Estimation Methods

Weighting was used in ALL to adjust for sampling and nonresponse. Responses to the literacy tasks were scored using item response theory (IRT) scaling. A multiple imputation procedure based on plausible values methodology was used to estimate the literacy proficiencies of individuals who completed literacy tasks.

Weighting. Each participating country in ALL used a multistage probability sample design with stratification and unequal probabilities of respondent selection. Furthermore, there was a need to compensate for the nonresponse that occurred at varying levels. Therefore, the estimation of population parameters and the associated standard errors was dependent on the survey weights. All participating countries used the same general procedure for calculating the survey weights. However, each country developed the survey weights according to its particular probability sample design. In general, two types of weights were calculated by each country: population weights that are required for the production of population estimates and jackknife replicate weights that are used to derive the corresponding standard errors.

Population weights. For each respondent record, the population weight was created first by calculating the theoretical or sample design weight, then by deriving a base sample weight by mathematically adjusting the theoretical weight for nonresponse. The base weight is the fundamental weight that can be used to produce population estimates. However, in order to ensure that the sample weights were consistent with a country's known population totals (i.e., benchmark totals) for key characteristics, the base sample weights were ratio-adjusted to the benchmark totals.

Jackknife weights. It was recommended that 10 to 30 jackknife replicate weights be developed for use in determining the standard errors of the survey estimates. Switzerland produced 15 jackknife replicate weights. The remaining countries produced 30 jackknife replicate weights.

Scaling. The results of ALL are reported along four scales—two literacy scales (prose and document), a single numeracy scale, and a scale capturing problem solving—with each ranging from 0 to 500 points. One might imagine these tasks arranged along their respective scale in terms of their difficulty for adults and the level of proficiency needed to respond correctly to each task. The procedure used in ALL to model these continua of difficulty and ability is IRT. IRT is a mathematical model used for estimating the probability that a particular person will respond correctly to a given task from a specified pool of tasks.

The scale value assigned to each item results from how representative samples of adults in participating countries perform on each item and is based on the theory that someone at a given point on the scale is equally proficient in all tasks at that point on the scale. For ALL, as for IALS, proficiency was determined to mean that someone at a particular point on the proficiency scale would have an 80 percent chance of answering items at that point correctly.

Just as adults within each participating country in ALL are sampled from the population of adults living in households, each task that was constructed and used in the assessment represents a type of task sampled from the domain or construct defined here. Hence, it is representative of a particular type of literacy, numeracy, or problem-solving task that is associated with adult contexts.

In an attempt to display the progression of complexity and difficulty from the lower end of each scale to the upper end, each proficiency scale was divided into levels. Both the literacy and numeracy scales used five levels, where Level 1 represents the lowest level of proficiency and Level 5 the highest. These levels are defined as follows: Level 1 (0 to 225), Level 2 (226 to 275), Level 3 (276 to 325), Level 4 (326 to 375), and Level 5 (376 to 500). The scale for problem solving used four levels, where Level 1 is the lowest level of proficiency and Level 4 the highest. These four levels are defined as follows: Level 1 (0 to 250), Level 2 (251 to 300), Level 3 (301 to 350), and Level 4 (351 to 500).

Since each level represents a progression of knowledge and skills, individuals within a particular level not only demonstrate the knowledge and skills associated with that level but the proficiencies associated with the lower levels as well. In practical terms, this means that individuals performing at 250 (the middle of Level 2 on one of the literacy or numeracy scales) are expected to be able to perform the average Level 1 and Level 2 tasks with a high degree of proficiency. A comparable point on the problem-solving scale would be 275. In ALL, as in IALS, a high degree of proficiency is defined in terms of a response probability of 80 percent. This means that individuals estimated to have a particular scale score are expected to perform tasks at that point on the scale correctly with an 80 percent probability. It also means they will have a greater than 80 percent chance of performing tasks that are lower on the scale. It does not mean, however, that individuals with given proficiencies can never succeed at tasks with higher difficulty values. It does suggest that the more difficult the task relative to their proficiency, the lower the likelihood of a correct response.

Imputation. A respondent had to complete the background questionnaire, correctly complete at least two out of six simple tasks from the core block of literacy tasks, and attempt at least five tasks per literacy scale in order for researchers to be able to estimate his or her literacy skills directly. Literacy proficiency data were imputed for individuals who failed or refused to perform the core literacy tasks and for those who passed the core block, but did not attempt at least five tasks per literacy scale. Because the model used to impute literacy estimates for nonrespondents relied on a full set of responses to the background questions, ALL countries were instructed to obtain at least a background questionnaire from sampled individuals. ALL countries were also given a detailed nonresponse classification to use in the survey.

Literacy proficiencies of respondents were estimated using a multiple imputation procedure based on plausible values methodology. Special procedures were used to impute missing cognitive data.

Literary proficiency estimation (plausible values). A multiple imputation procedure based on plausible

values methodology was used to estimate respondents' literacy proficiency in ALL. When a sampled individual decided to stop the assessment, the interviewer used a standardized nonresponse coding procedure to record the reason why the person was stopping. This information was used to classify nonrespondents into two groups: (1) those who stopped the assessment for literacy-related reasons (e.g., language difficulty, mental disability, or reading difficulty not related to a physical disability); and (2) those who stopped for reasons unrelated to literacy (e.g., physical disability or refusal). The reasons given most often by individuals for not completing the assessment were reasons related to their literacy skills; the other respondents gave no reason for stopping or gave reasons unrelated to their literacy.

When individuals cited a literacy-related reason for not completing the cognitive items, it implies that they were unable to respond to the items. On the other hand, citing reasons unrelated to literacy implies nothing about a person's literacy proficiency. Based on these interpretations, ALL adapted a procedure originally developed for the National Adult Literacy Survey to treat cases in which an individual responded to fewer than five items per literacy scale, as follows: (1) if the individual cited a literacy-related reason for not completing the assessment, then all consecutively missing responses at the end of the block of items were treated as wrong; and (2) if the individual cited reasons unrelated to literacy for not completing the assessment, then all consecutively missing responses at the end of a block were treated as "not reached."

Proficiency values were estimated based on respondents' answers to the background questions and the cognitive items. As an intermediate step, the functional relationship between these two sets of information was calculated, and this function was used to obtain unbiased proficiency estimates with reduced error variance. A respondent's proficiency was calculated from a posterior distribution that was the multiple of two functions: a conditional distribution of proficiency, given responses to the background questions; and a likelihood function of proficiency, given responses to the cognitive items.

Future Plans

The OECD plans to conduct another survey, the Program for the International Assessment for Adult Competencies (PIAAC). It is built on the knowledge and experiences gained from IALS and ALL. PIAAC will measure relationships between educational background, workplace experiences and skills, professional attainment, use of ICT, and cognitive skills in the areas of literacy, numeracy and problem-solving. The assessment will be administered to 5,000 adults from ages 16 to 65. Administration of the survey will occur in 2011, with results being released in early 2013.

5. DATA QUALITY AND COMPARABILITY

The literacy tasks contained in ALL and the adults asked to participate in the survey were samples drawn from their respective universes. As such, they were subject to some measurable degree of uncertainty. ALL implemented procedures to minimize both sampling and nonsampling errors. The ALL sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. Quality control activities were employed during interviewer training, data collection, and processing of the survey data.

Sampling Error

Because ALL employed probability sampling, the results were subject to sampling error. Although small, this error was higher in ALL than in most studies because the cost of surveying adults in their homes is so high. Most countries simply could not afford large sample sizes.

Each country provided a set of replicate weights for use in a jackknife variance estimation procedure.

Nonsampling Error

The key sources of nonsampling error in ALL were differential coverage across countries and nonresponse bias, which occurred when different groups of sampled individuals failed to participate in the survey. Other potential sources of nonsampling error included deviations from prescribed data collection procedures and errors of logic that resulted from mapping idiosyncratic national data into a rigid international format. Scoring error, associated with scoring open-ended tasks reliably within and between countries, also occurred. Finally, because ALL data were collected and processed independently by the various countries, the study was subject to uneven levels of commonplace data capture, data processing, and coding errors.

Coverage error. The design specifications for ALL stated that in each country the study should cover the civilian, noninstitutionalized population ages 16 to 65. It is the usual practice to exclude the institutionalized population from national surveys because of the difficulties in conducting interviews in institutional settings. Similarly, it is not uncommon to exclude

certain other parts of a country's population that pose difficult survey problems (e.g., persons living in sparsely populated areas). The intended coverage of the surveys generally conformed well to the design specifications: each of the ALL countries attained a high level of population coverage. However, it should be noted that actual coverage is generally lower than the intended coverage because of deficiencies in sampling frames and sampling frame construction (e.g., failures to list some households and some adults within listed households).

Nonresponse error. For ALL, several procedures were developed to reduce biases due to nonresponse, based on how much of the survey the respondent completed.

Unit nonresponse. The definition of a respondent for ALL was a person who partially or fully completed the background questionnaire. Unweighted response rates varied considerably from country to country, ranging from a high of 82 percent (Bermuda) to a low of 40 percent (Switzerland). The United States had an unweighted response rate of 66 percent (see table 17).

Several precautions were taken against nonresponse bias. Interviewers were specifically instructed to return several times to nonrespondent households in order to obtain as many responses as possible. In addition, all countries were asked to ensure that the address information provided to interviewers was as complete as possible in order to reduce potential household identification problems.

Table 17. Sample size and response rate for the United States for the Adult Literacy and Lifeskills Survey (ALL): 2003

Country	Population ages 16 to 65 (millions)	Initial sample size	Out-of-scope cases[1]	Number of respondents[2]	Unweighted response rate (percent)
United States	184	7,045	1,846	3,420	66

[1]Out-of-scope cases are those where the residents were not eligible for the survey, the dwelling could not be located, the dwelling was under construction, the dwelling was vacant or seasonal, or the cases were duplicates.

[2]A respondent's data are considered complete for the purposes of the scaling of a country's psychometric assessment data provided that at least the Background Questionnaire variables for age, gender, and education have been completed.

SOURCE: Desjardins, R., Murray, S., Clermont, Y., and Werquin, P. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey*. Ottawa, Canada: Statistics Canada.

Item nonresponse. Not-reached responses were classified into two groups: nonparticipation immediately or shortly after the background information was collected; and premature withdrawal from the assessment after a few cognitive items were attempted. The first type of not-reached response varied a great deal across countries according to the frames from which the samples were selected. The second type of not-reached response was due to quitting the assessment early, resulting in incomplete cognitive data. Not-reached items were treated as if they provided no information about the respondent's proficiency, so they were not included in the calculation of likelihood functions for individual respondents. Therefore, not-reached responses had no direct impact on the proficiency estimation for subpopulations. The impact of not-reached responses on the proficiency distributions was mediated through the subpopulation weights.

Measurement error. Assessment tasks were selected to ensure that, among population subgroups, each literacy domain (prose, document, numeracy, and problem solving) was well covered in terms of difficulty, stimuli type, and content domain. The ALL item pool was developed collectively by participating countries. Items were subjected to a detailed expert analysis at ETS and vetted by participating countries to ensure that the items were culturally appropriate and broadly representative of the population being tested. For each country, experts who were fluent in both English and the language of the test reviewed the items and identified ones that had been improperly adapted. Countries were asked to correct problems detected during this review process. To ensure that all of the final survey items had a high probability of functioning well, and to familiarize participants with the unusual operational requirements involved in data collection, each country was required to conduct a pilot survey.

Although the pilot surveys were small and typically were not based strictly on probability samples, the information they generated enabled ETS to reject items, to suggest modifications to a few items, and to choose good items for the final assessment. ETS's analysis of the pilot survey data and recommendations for final test design were presented to and approved by participating countries.

6. CONTACT INFORMATION

For content information on ALL, contact:

Eugene Owen
Phone: (202) 502-7422
E-mail: eugene.owen@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Desjardins, R., Murray, S., Clermont, Y., and Werquin, P. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey*. Ottawa, Canada: Statistics Canada.

Lemke, M., Miller, D., Johnston, J., Krenzke, T., Alvarez-Rojas, L., Kastberg, D., and Jocelyn, L. (2005). *Highlights From the 2003 International Adult Literacy and Lifeskills Survey (ALL)-(Revised)* (NCES 2005-117rev). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 25: Progress in International Reading Literacy Study (PIRLS)

1. OVERVIEW

The Progress in International Reading Literacy Study (PIRLS) is a large international comparative study of the reading literacy of fourth-grade students. The study is conducted by the International Association for the Evaluation of Educational Achievement (IEA), with national sponsors in each participating jurisdiction. The National Center for Education Statistics (NCES), in the Institute of Education Sciences at the U.S. Department of Education, is responsible for the implementation of PIRLS in the United States. Reading literacy is one of the most important abilities that students acquire as they progress through their early school years. It is the foundation for learning across all subjects, it can be used for recreation and for personal growth, and it equips young children with the ability to participate fully in their communities and the larger society. Participants in PIRLS include both countries and subnational entities, both of which are referred to as “jurisdictions.” PIRLS focuses on the achievement and reading experiences of children in grades equivalent to fourth grade in the United States. The study includes a written test of reading comprehension and a series of questionnaires focusing on the factors associated with the development of reading literacy. PIRLS was first administered in 2001 to students in 35 jurisdictions and was administered again in 2006 to students in 45 jurisdictions. The next PIRLS is scheduled for 2011.

PROGRESS IN INTERNATIONAL READING AND LITERACY STUDY:

Three aspects of reading literacy:

- Purpose for reading
- Processes of comprehension
- Reading behaviors and attitudes

Four sets of questionnaires:

- Student questionnaire
- Learning to read (home) survey
- Teacher questionnaire
- School principal questionnaire

Purpose

PIRLS is a carefully constructed reading assessment, consisting of a test of the reading literacy of fourth-grade students and questionnaires to collect information about fourth-grade students’ reading literacy performance. PIRLS has four goals: (1) develop internationally valid instruments for measuring reading literacy suitable for establishing internationally comparable literacy levels in each of the participating jurisdictions; (2) describe on one international scale the literacy profiles of fourth-graders in school in each of the participating jurisdictions; (3) describe the reading habits of fourth-graders in each participating jurisdiction; and (4) identify the home, school, and societal factors associated with the literacy levels and reading habits of fourth-graders in school.

Components

PIRLS focuses on three aspects of reading literacy: purposes for reading; processes of comprehension; student reading behaviors and engagement. The first two form the basis of the written test of reading comprehension. The student background questionnaire addresses the third aspect.

In PIRLS, purpose for reading refers to the two types of reading that account for most of the reading young students do, both in and out of school: (1) reading for literary experience, and (2) reading to acquire and use information. In the assessment, narrative fiction is used to assess students' ability to read for literary experience, while a variety of informational texts are used to assess students ability to acquire and use information while reading. The PIRLS assessment contains an equal proportion of texts assessing each purpose. Processes of comprehension refer

to ways in which readers construct meaning from the text. There are four comprehension processes: focusing on and retrieving specific ideas; making inferences; interpreting and integrating ideas and information; and examining or evaluating text features.

Assessment. The PIRLS assessment instruments include stories and informational texts at the fourth-grade level collected internationally. Students are asked to engage in a full repertoire of reading skills and strategies, including retrieving and focusing on specific ideas, making simple and more complex inferences, and examining and evaluating text features. The passages are followed by constructed-response and multiple-choice format questions about the text.

In PIRLS 2001, reading passages were printed in some students' assessment booklets, while other students were given the *PIRLS Reader*, a short anthology of a variety of reading texts, in addition to an assessment booklet. Using different booklets allows PIRLS to report results from more assessment items than can fit in one booklet, without making the assessment longer. To provide good coverage of each skill domain, the test items developed required over 5 hours of testing time. However, testing time was kept to 80 minutes for each student by clustering items in 8 blocks distributed across the 10 booklets, (9 student test booklets and the *PIRLS Reader*). Each student completed only one of the booklets. As a consequence, no student received all items, but each item was answered by a representative sample of students.

PIRLS 2006's design was built on PIRLS 2001. To evaluate changes in achievement over time, in 2006 new measuring scales were created in addition to the scale for reading achievement overall. To accommodate these changes, the booklet design expanded to include additional test booklets, and the total assessment time increased. PIRLS 2006 included 10 blocks, consisting of a reading passage and its accompanying questions. Four of the PIRLS 2001 test blocks were kept secure and carried forward for measuring trends in 2006, the six remaining blocks were redesigned. The new materials were added to reflect the broad approaches established for 2001, while refreshing and expanding the range of texts and devising items that brought out the qualities of each passage. The item blocks were then distributed across 13 booklets (including *PIRLS Reader*, a full color, magazine-style booklet) and each student was administered one of the booklets.

Questionnaires. Background questionnaires in PIRLS are administered to collect information about students' home and school experiences in learning to read. By gathering information about children's experiences (together with reading achievement on the PIRLS test), it is possible to identify the factors or combinations of factors that relate to high reading literacy. PIRLS 2001 and PIRLS 2006 administered questionnaires to students, teachers, and school principals. In jurisdictions other than the United States, a parent questionnaire is also administered. Additionally, PIRLS 2006 included a newly constructed curriculum questionnaire that provided information about the national context.

Student questionnaire. Each student taking the PIRLS reading assessment completes the student questionnaire. The questionnaire asks about aspects of students' home and school experiences, including instructional experiences and reading for homework, self-perceptions about and attitudes toward reading, out-of-school reading habits, computer use, home literacy resources, and basic demographic information, such as parents' educational level, language spoken at home, and student reading activities.

Learning to read (home) survey. The learning to read survey is completed by the parents or primary caregivers of each student taking the PIRLS reading assessment. It addresses child/parent literacy interactions, home literacy resources, parents' reading habits and attitudes, home/school connections, and basic demographic and socioeconomic indicators. This assessment was not administered in the United States in 2001 and 2006.

Teacher questionnaire. The reading teacher of each fourth-grade class sampled for PIRLS completes a questionnaire designed to gather information about classroom contexts for developing reading literacy. This questionnaire asks teachers about characteristics of the class tested (such as size, reading levels of the students, and language abilities of the students). It also asks about instructional time, materials and activities for teaching reading and promoting the development of students' reading literacy, and the grouping of students for reading instruction. Questions about classroom resources, assessment practices, and home/school connections are also included. The questionnaire also asks teachers for their views on opportunities for professional development and collaboration with other teachers and for information about their education and training.

School questionnaire. The principal of each school sampled for PIRLS responds to the school questionnaire. The questionnaire asks principals about enrollment and other school characteristics (such as where the school is located, resources available in the

surrounding area, and indicators of the socioeconomic background of the student body), characteristics of reading education in the school, instructional time, school resources (such as the availability of instructional materials and staff), home/school connections, and the school climate.

Curriculum questionnaire. First used in PIRLS 2006, this questionnaire focused on the nature of the development and implementation of a nationally (or regionally) defined reading curriculum in primary schools within each participating country.

In all, PIRLS takes 1½ to 2 hours of each student's time, including the assessment and background questionnaire.

In addition, system level information was provided by each participating country and published in the PIRLS 2001 Encyclopedia (Mullis et al. 2002) and the PIRLS 2006 Encyclopedia (Kennedy et al. 2007). The encyclopedias provide a description for each participating country of the policies and practices that guide school organization and classroom reading instruction in the lower grades.

Periodicity

PIRLS is administered once every 5 years, near the end of the school year in each jurisdiction. PIRLS was conducted in 2001 and 2006, and will be administered in the United States and other participating jurisdictions again in 2011.

2. USES OF DATA

PIRLS will help educators and policymakers by answering questions such as the following:

- How well do fourth-grade students read?
- How do students in one jurisdiction compare with students in another jurisdiction?
- Do fourth-grade students value and enjoy reading?
- Internationally, how do the reading habits and attitudes of students vary?

3. KEY CONCEPTS

International desired population. This is the grade or age level that each jurisdiction should address in its sampling activities. The international desired population for PIRLS 2001 was defined as all students enrolled in the upper of the two adjacent grades that contain the largest proportion of 9-year-olds at the time of testing. For PIRLS 2006, the international desired population was defined as all students enrolled in the grade that represents 4 years of schooling, counting from the 1st year of the International Standard Classification of Education (ISCED) Level 1, providing that the mean age at the time of testing was at least 9.5 years. For most jurisdictions, the target grade was the fourth grade or its national equivalent.

National desired population. PIRLS expects all participating jurisdictions to define their national desired population to correspond as closely as possible to the definition of the international desired population. For example, for PIRLS 2001, if the fourth grade was the upper of the two adjacent grades containing the greatest proportion of 9-year-olds in a particular jurisdiction, then students enrolled in fourth grade were the national desired population for that jurisdiction. For PIRLS 2006, if the fourth grade of primary school was the grade that represents 4 years of schooling in a particular jurisdiction (counting from the 1st year of ISCED Level 1), then students enrolled in fourth grade were the national desired population for that jurisdiction.

Although jurisdictions are expected to include all students in the target grade in their definition of the population, sometimes they have to reduce their coverage. Using its national desired population as a basis, each participating jurisdiction has to define its population in operational terms for sampling purposes. Ideally, the national defined population should coincide with the national desired population, although in reality there may be some school types or regions that cannot be included; consequently, the national defined population is usually a very large subset of the national desired population.

National Research Coordinators (NRCs) and data collection contractor. Each participating jurisdiction appoints a national research coordinator to monitor national data collection and processing in accordance with international standards. NCES contracts with a data collection firm to draw the samples, work with school coordinators, assemble and print the test booklets, and pack and ship the necessary materials to the sampled schools. The contractor is also

responsible for working with school coordinators, translating the test instruments, assembling and printing the test booklets, and packing and shipping the necessary materials to the sampled schools. They are also responsible for arranging the return of the testing materials from the school to the national center, preparing for and implementing the constructed-response scoring, entering the results into data files, conducting on-site quality assurance observations for a 10 percent sample of schools, and preparing a report on survey activities.

Reading literacy. The ability to use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential. This definition goes beyond simply decoding and comprehending text to include a broad range of information-processing skills that adults use in accomplishing the range of tasks associated with work, home, and community contexts. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers, and for enjoyment. In PIRLS, there is a distinction between reading for literary experience and reading to acquire and use information.

4. SURVEY DESIGN

Target Population

In IEA studies, the target population for all jurisdictions is known as the *international desired population*. The detailed definitions of international desired population for PIRLS 2001 and 2006 are provided in the section of Key Concepts. For both PIRLS 2001 and 2006, the international desired population corresponds to the fourth grade in most jurisdictions, including the United States. This population was chosen because it represents an important transition point in children's development as readers. In most jurisdictions, by the end of fourth grade, children are expected to have learned how to read, and are now reading to learn.

Sample Design

Using its national desired population as a basis, each participating jurisdiction has to define its population in operational terms for sampling purposes. PIRLS participants are expected to ensure that the national defined population includes at least 95 percent of the national desired population. Exclusions (which should be kept to a minimum) can occur at the school level, within the sampled schools, or at both levels. Because the national desired population is restricted to schools that contain the required grade, schools not containing the target grade are considered to be outside the scope of the sample—not part of the target population.

In each jurisdiction, representative samples of students are selected using a two-stage sampling design. In the first stage, at least 170 schools are selected using probability proportional to size (PPS) sampling. Jurisdictions can incorporate in their sampling design important reporting variables (for example, urbanicity or school type) as stratification variables. In the second stage, one or two fourth-grade classes are randomly sampled in each school. This results in a sample size of at least 3,750 students in each jurisdiction. Some jurisdictions opt to include more schools and classes, enabling additional analyses, which results in larger sample sizes. In 2006, PIRLS required that all student sample sizes should not be less than 4,000 students.

In the United States in 2001, a nationally representative sample of 3,760 fourth-grade students from 170 schools was selected. The schools were randomly selected first, and then one or two classrooms were randomly selected within each school. In the United States in 2006, a nationally representative sample of 5,190 fourth-grade students from 180 schools was selected. The schools were randomly selected first, and then one or two classrooms were randomly selected within each school.

First sampling stage. The sample selection method used for the first sampling stage in PIRLS makes use of a systematic PPS technique. In order to use this method, it is necessary to have some measure of size (MOS) of the sampling units. Ideally, this is the number of sampling elements within the units (e.g., the number of students in the school in the target grade). If this is unavailable, some other highly correlated measure, such as total school enrollment, is used. The schools in each explicit stratum are listed in order of the implicit stratification variables, together with the MOS for each school. Schools are further sorted by MOS within implicit stratification variables. The cumulative MOS is a measure of the size of the population of sampling elements; dividing it by the number of schools to be sampled gives the sampling interval.

The first school is sampled by choosing a random number in the range between 1 and the sampling interval. The school whose cumulative MOS contains the random number is the sampled school. By adding the sampling interval to that first random number, a second school is identified. This process of consistently adding the sampling interval to the previous selection number results in a PPS sample of the required size.

Very large jurisdictions have an opportunity to introduce a preliminary sampling stage before sampling schools. The Russian Federation and the United States avail themselves of this option. In these jurisdictions, the first step is to draw a sample of geographic regions using PPS sampling. Then a sample of schools is drawn from each sampled region. This design is used mostly as a cost reduction measure, where the construction of a comprehensive list of schools would have been either impossible or prohibitively expensive. Also, the additional sampling stage reduces the dispersion of the school sample, thereby potentially reducing travel costs. Sampling guidelines are put in place to ensure that an adequate number of units will be sampled from this preliminary stage.

Second sampling stage. The second sampling stage consists of selecting classrooms within sampled schools. As a rule, one classroom per school is sampled, although some participants opt to sample two classrooms. All classrooms are selected with equal probabilities for all jurisdictions. It is suggested that any classroom smaller than half the specified minimum cluster size be combined with another classroom from the same grade and school.

Trends in IEA's Reading Literacy Study. PIRLS jurisdictions that earlier participated in the 1991 IEA Reading Literacy Study had the option of undertaking the Trends in IEA's Reading Literacy Study, which measured trends in reading achievement using IEA's 1991 reading test and student questionnaire. Since the target population for the Trends in IEA's Reading Literacy Study is similar (but not identical) to the PIRLS target population, it is possible to use the PIRLS school sample as the basis for the trend study sample. Accordingly, the sampling plan for the Trends in IEA's Reading Literacy Study is simple: select every second school sampled for PIRLS, and from each of these, sample one additional classroom from the target grade. Since the sample of schools for the Trends in IEA's Reading Literacy Study is essentially a subsample of the PIRLS sample of schools, most of the required sampling tasks are carried out during the PIRLS school sampling.

Assessment Design

The PIRLS International Study Center is responsible for the design, development, and implementation of the study—including developing the instruments and survey procedures, ensuring quality in data collection, and analyzing and reporting the study results. The PIRLS Reading Development Group contributes to the framework and reading test. Committee members review various drafts of the framework and assessment blocks, and review and endorse the final reading test. The PIRLS Questionnaire Development Group, comprising representatives from nine countries, helps develop the PIRLS questionnaires (including writing items and reviewing drafts of all questionnaires).

Development of framework and questions. At the heart of the PIRLS assessment is the definition of reading literacy established by the Reading Development Group and refined by National Research Coordinators. The PIRLS definition of reading literacy builds on the definition used in the 1991 IEA study, but elaborates on that definition by making specific reference to reading by children.

In accordance with the framework, the passages in the reading test are authentic texts drawn from children's storybooks and informational sources. Submitted and reviewed by the PIRLS jurisdictions, the passages represent a range of types of literary and informational texts. The literary passages include realistic stories and traditional tales, while the informational texts include chronological and nonchronological articles, biographical articles, and informational leaflets.

Two item formats are used to assess children's reading literacy—multiple-choice and constructed-response. Each type of item is used to assess both reading purposes and all four reading processes.

Matrix sampling. PIRLS has ambitious goals for covering the domain of reading literacy. The Reading Development Group felt that at least eight passages and items (four for each reading purpose) were needed to provide a valid and reliable measure of reading achievement. Since it would not be possible to administer the entire test to any one student, PIRLS used a matrix sampling technique to distribute the assessment material among students, yet retain linkages necessary for scaling the achievement data.

In PIRLS 2001, assessment material was divided into 40-minute "blocks," each comprised of a passage (a story or article) and items representing at least 15 score points. There were eight such blocks, four for each reading purpose. The eight assessment blocks were distributed across 10 test booklets, and each student completed one booklet in an 80-minute testing session. Each booklet contained two blocks—two literary, two informational, or one of each—and most blocks appeared in three booklets. One of the 10 booklets was the *PIRLS Reader*, a color booklet containing two reading passages; the test items for it were located in a separate booklet. The two blocks for the *Reader* appeared only in that booklet. The distribution of blocks across booklets "links" the booklets to enable

the achievement data to be scaled using Item Response Theory (IRT) methods.

The new material developed for PIRLS 2006 was combined with the four secure blocks retained from the 2001 assessment, providing an overall assessment that would allow the calculation of trends over 5 years. The PIRLS 2006 reading assessment was comprised of 13 booklets, one of which was administered to each student. Each booklet contained two blocks, comprised of a story or article followed by a series of questions pertaining to the text passage. In 2006, there were 10 blocks in total (5 for each reading purpose), which were systematically rotated throughout the booklets. As in 2001, the two blocks for the *Reader* appeared only in that booklet.

Data Collection and Processing

Reference dates. PIRLS is administered near the end of the school year in each jurisdiction. For PIRLS 2001, in jurisdictions in the Northern Hemisphere (where the school year typically ends in May or June), the assessment was conducted in April, May, or June 2001.

In the PIRLS 2006, jurisdictions in the Northern Hemisphere conducted the assessment between March and May 2006. In the United States, data collection began slightly earlier and ended in early June. In the Southern Hemisphere, the school year typically ends in November or December; in these jurisdictions, the assessment was conducted in October or November in 2001 and in October and November in 2005.

Data collection. Each jurisdiction is responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Manuals provide explicit instructions to the NRCs and their staff members on all aspects of the data collection—from contacting sampled schools to packing and shipping materials to the IEA Data Processing Center for processing and verification. Manuals are also prepared for test administrators and for individuals in the sampled schools who work with the national centers to arrange for the data collection within the schools. These manuals address all aspects of the assessment administration within schools (including test security, distribution of booklets, timing and conduct of the testing session, and returning materials to the national center).

The PIRLS International Study Center places great emphasis on monitoring the quality of the PIRLS data collection. In particular, the Study Center implements an international program of site visits, whereby international Quality Control Monitors (QCMs) visit a sample of 15 schools in each jurisdiction and observe the test administration. In addition to the international program, NRCs are also expected to organize an independent national quality control program based upon the international model. The latter program requires national QCMs to document data collection activities in their jurisdiction. The national QCMs visit a random sample of 10 percent of the schools (in addition to those visited by the international QCMs) and monitor the testing sessions—recording their observations for later analysis.

Editing. To ensure the availability of comparable, high-quality data for analysis, PIRLS takes rigorous quality control steps to create the international database. PIRLS prepares manuals and software for jurisdictions to use in creating and checking their data files, so that the information will be in a standardized international format before being forwarded to the IEA Data Processing Center (DPC) in Hamburg for creation of the international database. Upon arrival at the DPC, the data undergo an exhaustive cleaning process involving several iterative steps and procedures designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. The process also emphasizes consistency of information within national datasets and appropriate linking among the student, parent, teacher, and school data files.

Throughout the process, the data are checked and double-checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers are contacted regularly and given multiple opportunities to review the data for their jurisdictions. In conjunction with the IEA Data Processing Center, the International Study Center reviews item statistics for each cognitive item in each jurisdiction to identify poorly performing items. In general, the items exhibit very good psychometric properties in all jurisdictions.

Estimation Methods

Weighting. Sampling weights are calculated according to a three-step procedure involving selection probabilities for schools, classrooms, and students.

School weight. The first step consists of calculating a school weight, which also incorporates weighting factors from any additional front-end sampling stages, such as districts or regions. A school-level participation adjustment is then made to the school weight to compensate for any sampled schools that do not participate. This adjustment is calculated independently for each explicit stratum.

The PIRLS sample design requires that school selection probabilities be proportional to the school size, defined as enrollment in the target grade. For jurisdictions with a preliminary sampling stage (such as the United States and the Russian Federation), the basic first-stage weight also incorporates the probability of selection in this preliminary stage. The first-stage weight in such cases is simply the product of the "region" weight and the first-stage weight.

In some jurisdictions, schools are selected with equal probabilities. This generally occurs when a large sampling ratio is used. Also, in some jurisdictions, explicit or implicit strata are defined to deal with very large schools or small schools. Equal probability sampling is necessary in these strata.

First-stage weights are calculated for all sampled and replacement schools that participate. A school-level participation adjustment is required to compensate for those schools that are sampled but do not participate and, hence, are not replaced. Sampled schools that are found to be ineligible are removed from the calculation of this adjustment. The school-level participation adjustment is calculated separately for each explicit stratum.

Classroom weight. In the second step, a classroom weight reflecting the probability of the sampled classroom(s) being selected from all the classrooms in the school at the target grade level is calculated. All classrooms are sampled with equal probability. No classroom-level participation adjustment is necessary, since in most cases a single classroom is sampled in each school. If a school agrees to take part in the study, but the classroom refuses to participate, adjustment for nonparticipation is made at the school level. If one of two selected classrooms in a school does not participate, then the classroom weight is calculated as though a single classroom has been selected in the first place. The classroom weight is calculated independently for each school.

Student weight. Because intact classrooms are sampled in PIRLS, each student in the sampled classrooms is certain of selection, so the base student weight is 1.0. However, as a third and final step, a nonparticipation adjustment is made to compensate for students who do not take part in the testing. This is calculated independently for each sampled classroom. The basic sampling weight attached to each student record is the product of the three intermediate weights: the first-stage (school) weight, the second-stage (classroom) weight, and the third-stage (student) weight.

Overall sampling weight. The overall student sampling weight is the product of the three weights, including the nonparticipation adjustments.

Scaling. The primary approach to reporting PIRLS achievement data is based on IRT scaling methods. The IRT analysis provides a common scale on which performance can be compared across countries. Student reading achievement is summarized using a family of IRT models. In 2006 PIRLS, 2- and 3-parameter logistic IRT models were used for dichotomously scored items, and generalized partial credit models for constructed-response items with two or three available score points. The IRT methodology is preferred for developing comparable estimates of performance for all students, since students respond to different passages and items depending upon which of the test booklets they receive. This methodology produces a score by averaging the responses of each student to the items that he or she takes in a way that takes into account the difficulty and discriminating power of each item. The approach followed in PIRLS uses information from the background questionnaires to provide improved estimates of student performance (a process known as conditioning) and multiple imputation to generate student scores (or "plausible values") for analysis and reporting.

In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within jurisdictions vary and provide information on percentiles of performance. Treating all participating jurisdictions equally, the PIRLS scale average across jurisdictions was set to 500 and the standard deviation to 100. Since the jurisdictions vary in size, each jurisdiction is weighted to contribute equally to the mean and standard deviation of the scale. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation.

In the PIRLS 2001 analysis, achievement scales were produced for each of the two reading purposes—reading for literary experience and reading for information—as well as for reading overall. The PIRLS 2006 reading achievement scales were designed to provide reliable measures of student achievement common to both the 2001 and 2006 assessments, based on the metric established originally in 2001. In 2006 PIRLS, in addition to the scale for reading achievement overall, IRT scales were created to measure changes in achievement in the two purposes of reading and two overarching reading processes.

Imputation. No imputations are generated for missing values. However, multiple imputations are used to

generate student scores (or "plausible values") for analysis and reporting.

The PIRLS item pool is far too extensive to be administered in its entirety to any one student, and so a matrix-sampling test design was developed whereby each student is given a single test booklet containing only a part of the entire assessment. The results for all of the booklets are then aggregated using IRT techniques to provide results for the entire assessment. Since each student responds to a subset of the assessment items, multiple imputations (the generation of "plausible values") are used to derive reliable estimates of student performance on the assessment as a whole. Since every student proficiency estimate incorporates some uncertainty, PIRLS follows the customary procedure of generating five estimates for each student and using the variability among them as a measure of this imputation uncertainty, or error. In the PIRLS international reports (Mullis et al. 2003, 2007), the imputation error for each variable is combined with the sampling error for that variable to provide a standard error incorporating both.

5. DATA QUALITY AND COMPARABILITY

A group of distinguished international reading scholars, the Reading Development Group, was formed to construct the PIRLS framework and endorse the final reading assessment. Each jurisdiction followed internationally prescribed procedures to ensure valid translations and representative samples of students. The national QCMs compared the final version of the booklets with the international translation verifier's comments to ensure that their suggestions had been incorporated appropriately into the materials. The QCMs were then appointed in each jurisdiction to monitor the testing sessions at the schools to ensure that the high standards of the PIRLS data collection process were met.

Sampling Error

The standard errors of the reading proficiency statistics reported by PIRLS include both sampling and imputation variance components.

When, as in PIRLS, the sampling design involves multistage cluster sampling, there are several options for estimating sampling errors that avoid the assumption of simple random sampling. The jackknife repeated replication technique (JRR) is chosen by PIRLS because it is computationally straightforward and provides approximately unbiased estimates of the sampling errors of means, totals, and percentages.

The particular application of the JRR technique used in PIRLS is termed a paired selection model because it assumes that the primary sampling units (PSUs) can be paired in a manner consistent with the sample design, with each pair regarded as members of a pseudo-stratum for variance estimation purposes. When used in this way, the JRR technique appropriately accounts for the combined effect of the between- and within-PSU contributions to the sampling variance. The general use of JRR entails systematically assigning pairs of schools to sampling zones, and randomly selecting one of these schools to have its contribution doubled and the other to have its contribution zeroed, so as to construct a number of "pseudo-replicates" of the original sample. The statistic of interest is computed once for the original sample, and once again for each pseudo-replicate sample. The variation between the estimates for each of the replicate samples and the original sample estimate is the jackknife estimate of the sampling error of the statistic.

To apply the JRR technique used in PIRLS 2001 and PIRLS 2006, the sampled schools were paired and assigned to a series of groups known as "sampling zones." In total, 75 zones were used, allowing for 150 schools per jurisdiction. When more than 75 zones were constructed, they were collapsed to keep the total number to 75. For more information on sampling error, see the PIRLS technical reports (Martin, Mullis, and Kennedy 2003, 2007).

Imputation error. For each of the PIRLS reading scales, reading overall, and literary and informational reading, the IRT scaling procedure yields five imputed scores or plausible values for every student. The difference between the five values reflects the degree of uncertainty in the imputation process.

The general procedure for estimating the imputation variance using plausible values is the following. First compute the statistic (t) for each set of plausible values (M). The statistic t_m, where m = 1, 2, …, 5, can be anything estimable from the data, such as a mean, the difference between means, percentiles, and so forth. Once the statistics are computed, the imputation variance is then computed as

$$Var_{imp} = (1 + 1/M)Var(t_m)$$

where M is the number of plausible values used in the calculation, and $Var(t_m)$ is the variance of the estimates computed using each plausible value.

Nonsampling Error

Due to the particular situations of individual PIRLS jurisdictions, sampling and coverage practices have to be adaptable in order to ensure an internationally comparable population. As a result, nonsampling errors in PIRLS can be related both to coverage error and nonresponse.

Coverage error. PIRLS expects all participating jurisdictions to define their national desired population to correspond as closely as possible to its definition of the international desired population. Although jurisdictions are expected to include all students in the target grade in their definition of the population, sometimes they have to reduce their coverage. Although jurisdictions were expected to do everything possible to maximize coverage of the population by the sampling plan, schools could be excluded if they were in geographically remote regions, if they were of extremely small size, if they offered a curriculum or a school structure that was different from that found in the mainstream education system, or if they provided instruction only to students in the categories defined as "within-school exclusions."

Table 18. Weighted U.S. response rates for 2001 and 2006 PIRLS assessments

Year	School response rate	Student response rate	Overall response rate
2001	86	96	83
2006	86	95	82

NOTE: All weighted response rates refer to final adjusted weights. Response rates were calculated using the formula developed by the IEA for PIRLS. The standard NCES formula for computing response rates would result in a lower school response rate. Response rates are after replacement.
SOURCE: Martin, M.O., Mullis, I.V.S., and Kennedy, A.M. (Eds.). (2003). *PIRLS 2001 Technical Report*. Boston College, International Study Center. Chestnut Hill, MA. Baer, J., Baldi, S., Ayotte, K., and Green, P. (2007). *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading and Literacy Study (PIRLS)* (NCES 2008-017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Within-school exclusions were limited to students who, because of some disability, were unable to take the PIRLS tests, including educable mentally disabled students, functionally disabled students, and non-native-language speakers.

Nonresponse error.

Unit nonresponse. Unit nonresponse error results from nonparticipation of schools and students. Weighted and unweighted school and student response rates for PIRLS are computed for each participating jurisdiction. To monitor school participation, three school participation rates are computed: one using only originally sampled schools; one using sampled and first replacement schools; and one using sampled and both first and second replacement schools. Student participation rates are also computed, as are overall participation rates.

The minimum acceptable school-level response rate, before the use of replacement schools, was set at 85 percent. Likewise, the minimum acceptable student-level response rate was set at 85 percent. Jurisdictions understood that the goal for sampling participation was 100 percent of all sampled schools and students. Guidelines for reporting achievement data for jurisdictions securing less than full participation were modeled after IEA's Trends in International Mathematics and Science Study (TIMSS). Jurisdictions were assigned to one of three categories on the basis of their sampling participation. Jurisdictions in Category 1 were considered to have met the PIRLS sampling requirements and to have an acceptable participation rate. Jurisdictions in Category 2 met the sampling requirements only after including replacement schools. Jurisdictions that failed to meet the participation requirements, even with the use of replacement schools, were assigned to Category 3.

In 2001, almost all jurisdictions met the PIRLS sampling requirements and belonged in Category 1. Because they met the sampling requirements only after including replacement schools, England, the Netherlands, and the United States belonged in Category 2. Although Morocco and Scotland had overall weighted participation rates of 69 and 74 percent, respectively (even after including replacement schools), it was decided that these rates did not warrant the placement of the jurisdictions in Category 3. Instead, the results for Morocco and Scotland were annotated to indicate that they nearly satisfied the guidelines for sample participation rates after including replacement schools.

In 2006, almost all jurisdictions met the PIRLS sampling requirements and belonged in Category 1. Because they met the sampling requirements only after including replacement schools, Scotland, the United States, the Netherlands, and Belgium (Flemish) were placed in Category 2. Although Norway had overall participation rates after including replacement schools of just below 75 percent (71 percent), it was

decidedduring the sampling adjudication that this rate did not warrant placement in Category 3. Instead, the results for Norway were annotated in the 2006 international report similarly to what was done for Morocco and Scotland in 2001.

Data Comparability

IEA Reading Literacy Study and PIRLS. In 1991, the IEA launched the Reading Literacy Study, which assessed the reading literacy of 4th- and 9th-grade students in 32 jurisdictions. In 2001, IEA launched PIRLS in 35 jurisdictions. Although built on the foundation of the 1991 study, PIRLS is a new and different study, with a new assessment framework describing the interaction between two major reading purposes (literary and informative) and a range of four comprehension processes, an innovative reading test, and newly developed questionnaires for parents, students, teachers, and school principals.

Because the PIRLS 2001 reading test differed in a number of respects from the 1991 test, it was not possible to link the results of the two studies directly together. However, since PIRLS 2001 was scheduled to collect data on fourth-grade students 10 years after the 1991 Reading Literacy Study, PIRLS jurisdictions that participated in 1991 were given the opportunity of measuring changes in reading literacy achievement over that period by re-administering the 1991 reading literacy test to primary and elementary school students as part of the PIRLS data collection. The resulting study is known as the Trends in IEA's Reading Literacy Study. In 2001, nine jurisdictions replicated the 1991 Reading Literacy Study: Greece, Hungary, Iceland, Italy, New Zealand, Singapore, Slovenia, Sweden, and the United States. Conducted at the third or fourth grades (the grade with the most 9-year-olds), the study assessed student reading in three major domains: narrative texts, expository texts, and documents. Students completed a brief questionnaire about their home and school literacy activities and instruction. For more information on the trend study, see *Trends in Children's Reading Literacy Achievement 1991–2001: IEA's Repeat in Nine Countries of the 1991 Reading Literacy Study* (Martin et al. 2003). No such trend study was administered in conjunction with the 2006 PIRLS.

The United States conducted a study to compare the two international studies in the aspects of reading literacy each assessed, the types of texts they used in the assessments, and the types and difficulty of the questions they used. Both differences and similarities were found. The definitions of reading literacy were very similar. The types of passages used were similar, but in actually choosing and categorizing passages, the Reading Literacy Study emphasized the types of texts, while PIRLS focused on purposes for reading. In most cases, the passages and texts in PIRLS were longer, more engaging, and more complex. The question taxonomies that were generated to frame the tasks in the assessments were very different. The Reading Literacy Study taxonomy had a text focus with activities such as verbatim responses, main theme, and locating information. The PIRLS taxonomy suggested more consideration of the readers' interaction with the passage, especially in the categories of "interpret and integrate ideas and information" and "examine and evaluate content, language, and textual elements." The use of a high number of constructed-response items permitted the PIRLS questions to tap a wider range of reading responses; this is supported by the limited analysis of a sample of questions using Wixso's Levels of Depth of Knowledge. In general, PIRLS called for a wider range of skills than did the Reading Literacy Study, especially skills requiring deeper thinking. Also, the PIRLS passages were presented in an engaging and authentic manner that might have improved students' motivation to read and respond to the texts. This is one area where the form of PIRLS might have contributed to students' level of performance. However, if students lacked the skills necessary to respond to the items, engaging texts would not have helped much. For more information on the comparison study, see the *PIRLS-IEA Reading Literacy Framework: Comparative Analysis of the 1991 IEA Reading Study and the Progress in International Reading Literacy Study* (Kapinus 2003).

National Assessment of Educational Progress (NAEP) and PIRLS. To date, there have been two studies undertaken to compare the frameworks, reading passages, and assessment items of NAEP and PIRLS. The first study compared NAEP 2002 and PIRLS 2001 at both the framework and item levels. The second study updates with analysis of the passages and item sets added in NAEP 2007 and PIRLS 2006.

Definitions and organizations. In terms of how the domain is defined, there is considerable overlap between the NAEP and PIRLS concepts of reading literacy. The differences are relatively minor: the PIRLS framework is more explicit about its targeting to young readers and acknowledges a more diverse set of reading contexts such as for personal enjoyment (versus the NAEP framework, which focuses more on school-based reading and is intended to be generally applicable across younger to older grades).

In terms of the organization of the frameworks, both NAEP and PIRLS are organized around two dimensional matrices, which specify processes (i.e., the

cognitive element) and the purposes or contexts for which students read. In particular, there are some notable differences at the framework level in how the processes (called aspects in NAEP) are broken out and elaborated. NAEP's four categories include: forming a general understanding, developing an interpretation, making reader-text connections, and examining content and structure. PIRLS' four categories include: locating and retrieving explicitly stated information, making straightforward inferences, interpreting and integrating ideas and information, and examining and evaluating content, language and textual elements. The key areas of difference are that there is no apparent counterpart in the NAEP framework to the PIRLS locate and retrieve category, and there is no explicit counterpart in the PIRLS framework to the NAEP category that requires readers to think beyond the text and apply it to the real world (i.e., make reader-text connections).

In terms of the purposes for which students read, both frameworks specify a literary purpose and an information-related purpose. While the literary purposes seem to be defined in a similar way across the assessments, the information-related purposes suggest slight differences. PIRLS assesses not just reading to acquire information, but also to use information, in a way that goes beyond NAEP's definition. At the older grades, the NAEP framework includes a "reading to perform a task" purpose, which focuses on reading to learn how to do something, which is more similar to the use information aspect of PIRLS' "reading to acquire and use information purpose.

Passage and item analyses. The types of passages included in NAEP and PIRLS reflect the purposes that are assessed. In NAEP, students are presented with short stories, legends, biographies, and folktales, as well as magazine articles that focus on people, places, and events of interest to children—to cover both its literary experience and information purposes. Similarly, PIRLS also presents narrative fiction, usually in the form of short stories, as well as informational articles and, distinct from NAEP, brochures to cover its two similar purposes. Both NAEP and PIRLS strive to be "authentic" in that they try to present passages and items that would be encountered in and out of school. NAEP specifically calls for the use of authentic texts, and all passages are shown as previously published and generally are not edited at all (in terms of content or formatting) for use in NAEP. PIRLS also strives to use previously published texts, but has a more liberal policy on editing and changing the format of the texts used—which is sometimes necessary in an international context in order to meet constraints of translation to multiple languages and for culturally diverse participants. U.S. experts who have examined the PIRLS passages have noted the more edited, and sometimes less continuous, nature of some of these than the NAEP passages, particularly among passages for information purpose.

Altogether, the NAEP and PIRLS fourth-grade assessments each include 10 reading passages, although each student receives only a subset of those passages. In terms of length, the PIRLS passages tend to be shorter than the NAEP passages, averaging 707 words per passage compared to NAEP's 823 words per passage. The PIRLS passages range from 403 to 855 words; NAEP passages range from 644 to 1,361 words.

Readability analyses also suggest that the PIRLS passages may be slightly easier than NAEP. On a very simple measure, for example, sentence counts show that the PIRLS passages, with a higher number of sentences per 100 word sample, consist of shorter sentences on average than do the NAEP passages. On other more elaborate measures, such as Fry and Flesch analyses, which use sentence count along with syllable count to determine a corresponding age and grade level for each text, PIRLS passages are calculated to be about one grade level below the NAEP passages. Finally, a Lexile measure, which indicates the reading demand of the text in terms of semantic difficulty (vocabulary) and syntactic complexity (sentence length) and which is more recently developed and normed than the other measures, also suggests that the PIRLS passages are suitable for one to two grades below those from NAEP. It should be noted, however, that both assessments do include a range of passages suited below and above the targeted grade level to capture the range of reading ability.

Each of these passages has items associated with it—approximately 12-13 per passage in PIRLS and 10 per passage in NAEP. The two assessments are similar in that the majority of items on both assessments require students to develop an interpretation about what they have read, although there is a greater emphasis on this in NAEP, with 69 percent of items classified as such compared to 60 percent of the PIRLS items. PIRLS also has a notably smaller percentage of items classified as forming a general understanding or making reader text connections, having half or less the percentage NAEP has in those categories. One of the major differences between the two assessments, however, is that there are a number of PIRLS items (21 percent) that do not fit on the NAEP framework at all. In nearly all cases, these are items that ask the reader to retrieve explicitly stated information, which is not a skill delineated in the NAEP framework or found in its items.

For more information on the similarities and differences between PIRLS and NAEP, see *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments* (Binkley and Kelly 2003), and *Comparing PIRLS and PISA with NAEP in Reading, Mathematics, and Science* (Stephens, and Coleman, 2007).

6. CONTACT INFORMATION

For content information about PIRLS, contact:

Stephen Provasnik
Phone: (202) 502-7480
E-mail: stephen.provasnik@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

Most of the technical documentation for PIRLS is published by the International Study Center at Boston College. The U.S. Department of Education, National Center for Education Statistics, is the source of several additional references listed below; these publications are indicated by an NCES number**.**

General

Kennedy, A.M., Mullis, I.V.S., Martin, M.O., and Trong, K.L. (2007). *PIRLS 2006 Encyclopedia: A Guide to Reading Education in the Forty PIRLS 2006 Countries*. Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., and Kennedy, A.M. (2003). *Trends in Children's Reading Literacy Achievement 1991–2001: IEA's Repeat in Nine Countries of the 1991 Reading Literacy Study*. Boston College, International Study Center. Chestnut Hill, MA.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Kennedy, A.M. (Eds.). (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools in 35 Countries*. Boston College, International Study Center. Chestnut Hill, MA.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., and Flaherty, C.L. (2002). *PIRLS 2001 Encyclopedia: A Reference Guide to Reading Education in the Countries Participating in IEA's Progress in International Reading Literacy Study (PIRLS).* Boston College, International Study Center. Chestnut Hill, MA.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., and Foy, P. (2007). *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in Primary Schools in 40 Countries*. Boston College, International Study Center. Chestnut Hill, MA.

Uses of Data

Foy, P., and Kennedy, A.M. (2008). *PIRLS 2006 User Guide for the International Database.* Boston College, International Study Center. Chestnut Hill, MA.

Gonzalez, E.J., and Kennedy, A.M. (2003). *PIRLS 2001 User Guide for the International Database.* Boston College, International Study Center. Chestnut Hill, MA.

Survey Design

Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., and Sainsburg, M. (2001). *Framework and Specifications for PIRLS Assessment 2001—2nd Edition*. Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., Mullis, I.V.S., and Kennedy, A.M. (Eds.). (2003). *PIRLS 2001 Technical Report.* Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., Mullis, I.V.S., and Kennedy, A.M. (Eds.). (2007). *PIRLS 2006 Technical Report.* Boston College, International Study Center. Chestnut Hill, MA.

Mullis, I.V.S., Kennedy, A.M., Martin, M.O., and Sainsburg, M. (2006). *PIRLS 2006 Assessment Frameworks, 2nd Edition.* Boston College, International Study Center. Chestnut Hill, MA.

Ogle, L.T., Sen, A., Pahlke, E., Jocelyn, L., Kastberg, D., Roey, S., and Williams, T. (2003). *International Comparisons in Fourth-Grade Reading Literacy: Findings From the Progress in International Reading Literacy Study of 2001* (NCES 2003-073). National Center for Education Statistics, Institute of

Education Sciences, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Baer, J., Baldi, S., Ayotte, K., and Green, P. (2007). *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading and Literacy Study (PIRLS)* (NCES 2008-017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Binkley, M., and Kelly, D.L. (2003). *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments* (NCES Working Paper 2003-10). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kapinus, B. (2003). *PIRLS-IEA Reading Literacy Framework: Comparative Analysis of the 1991 IEA Reading Study and the Progress in International Reading Literacy Study* (NCES Working Paper 2003-05). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Piesse, A., and Rust, K. (2003). *U.S. 2001 U.S. PIRLS Nonresponse Bias Analysis* (NCES 2003-21). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Stephens, M., and Coleman, M. (2007). Comparing PIRLS and PISA with NAEP in Reading, Mathematics, and Science (Working Paper). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Available at:http://nces.ed.gov/Surveys/PISA/pdf/comppaper12082004.pdf.

Chapter 26: National Household Education Surveys Program (NHES)

1. OVERVIEW

BIENNIAL SAMPLE SURVEY OF HOUSEHOLD MEMBERS

NHES addresses topical issues on a rotating basis:

- Adult education and lifelong learning
- Before- and after-school programs and activities
- Early childhood education and school readiness
- Parent/family involvement in education

The National Household Education Surveys Program (NHES) conducts telephone surveys of the noninstitutionalized, civilian population of the United States. These surveys are designed to provide information on educational issues that are best addressed by contacting households rather than schools or other education institutions. They offer policymakers, researchers, and educators a variety of statistics on the condition of education in the United States.

Purpose

To (1) provide reliable estimates of the U.S. population regarding specific education-related topics; and (2) conduct repeated measurements of the same educational phenomena at different points in time.

Components

The NHES program for a given year typically consists of (1) a screener (an interview that collects household composition and demographic data); and (2) two or three surveys (extended interviews addressing specific education-related topics). However, in 1999, the surveys collected information on key indicators from the broad range of topics addressed in previous NHES survey cycles.

Adult Education. Surveys on this topic were administrated in 2005, 2003, 2001, 1999, 1995, and 1991.

The 2005 Adult Education Survey (AE-NHES:2005) collected data about participation in the following types of formal adult education activities: English as a Second Language (ESL), basic skills and high school completion, postsecondary degree and diploma programs, apprenticeships, work-related courses, and personal interest courses. Information on a new topic, informal learning activities for personal interest, was gathered as well.

The 2003 Adult Education for Work-Related Reasons Survey (AEWR-NHES:2003) collected information about participation in college and university degree or certificate programs taken for work-related reasons, postsecondary degree programs taken for work-related reasons, apprenticeships, work-related courses, and work-related informal learning. Additionally, the survey explored factors associated with participation or nonparticipation in adult education activities.

The Adult Education and Lifelong Learning Survey (AELL-NHES:2001) was administered in 2001. It collected data on type of program, employer support, and credential sought for participation in the following types of adult education activities: ESL, adult basic education, credential programs, apprenticeships, work-related. courses, and personal interest courses. Some information on informal learning activities at work was gathered as well.

In 1999, the Adult Education Survey (AE-NHES:1999) included questions on education background, work experience, participation in adult education (including participation through distance learning), literacy activities, community involvement, adult demographic characteristics, and household characteristics. Eligible respondents were 16 years of age or older who were not currently enrolled in 12th grade or below and not institutionalized or on active duty in the U.S. Armed Forces.

AE-NHES:1995 included questions concerning respondents' participation in basic skills courses, ESL courses, credential (degree or diploma) programs, apprenticeships, work-related courses, personal development/interest courses, and interactive video or computer training on the job. Information collected on programs and courses included the subject matter, duration, cost, location and sponsorship, and employer support. Nonparticipants in selected types of adult education were asked about their interest in educational activities and barriers to participation. Extensive background, employment, and household information were collected for each adult. Eligible respondents included civilians age 16 and older not currently enrolled in secondary school.

In AE-NHES:1991, eligible respondents were persons 16 years of age or older, identified as having participated in an adult education activity in the previous 12 months. The information collected on programs and up to four courses included the subject matter, duration, sponsorship, purpose, and cost. A smaller sample of nonparticipants in adult education also completed interviews about barriers to participation. Information on the household and the adult's background and current employment was also collected in this survey.

Before- and After-School Programs and Activities.
The Before- and After-School Programs and Activities Survey, conducted in 2005 and 2001 (ASPA-NHES:2005 and ASPA-NHES:2001), collected detailed information from parents of 9,580 children in kindergarten through eighth grade about the before- and after-school arrangements in which their children participated, including care by relatives or nonrelatives in private homes, before- or after-school programs in centers and in schools, activities that might provide adult supervision in the out-of-school hours, and children's self-care. Items also addressed continuity of care arrangements, parental perceptions of quality, reasons for choosing parental care, and obstacles to participation in nonparental arrangements. Information was also collected on children's health and disability status and on characteristics of the parents and household.

Civic Involvement. Civic Involvement Surveys were administered in 1999 and 1996. The 1999 Youth Survey (Youth-NHES:1999) expanded on one of the 1996 surveys: the 1996 Youth Civic Involvement Survey (YCI-NHES:1996). It included questions on the school learning environment, family learning environment, plans for future education, participation in activities that promote or indicate personal responsibility, participation in community service or volunteer activities, exposure to information about politics and national issues, political attitudes and knowledge, skills related to civic participation, and type and purpose of community service. A subset of youth who reported participation in community service were asked additional questions about their service experiences. Eligible respondents were youth in the grades 6 through 12.

Three Civic Involvement Surveys were conducted in 1996: the Parent and Family Involvement in Education/Civic Involvement Survey (PFI/CI-NHES:1996), the Youth Civic Involvement Survey (YCI-NHES:1996), and the Adult Civic Involvement Survey (ACI-NHES:1996). They included questions on sources of political information, civic participation, and knowledge and attitudes about government. YCI-NHES:1996 also provided an assessment of the opportunities that youth have to develop the personal responsibility and skills that would facilitate their taking an active role in civic life. Eligible respondents were (1) parents of students in grades 6 through 12 (including homeschooled students in those grades), (2) youth in grades 6 through 12, and (3) adults.

Early Childhood Education and School Readiness. Early Childhood Education Surveys were conducted in 2005, 2001, 1995, and 1991, and a School Readiness Survey was conducted in 2007 and 1993.

The Early Childhood Program Participation Survey of 2005 (ECPP-NHES:2005) was the fifth collection for this topic. It provided data on the early childhood program participation of infants, toddlers, and preschoolers as well as the ability to measure change over time. It gathered information on the nonparental care arrangements and education programs of preschool children, consisting of care by relatives; care by persons to whom the children were not related; and participation in day care centers and preschool programs, including Head Start. Eligible respondents to ECPP Surveys were the parents of children between birth and 3rd grade. The interview was conducted with

the parent most knowledgeable about the child's education or care.

ECPP-NHES:2001 gathered information on the nonparental care arrangements and education programs of preschool children, which included care by relatives; care by persons to whom the children were not related; and participation in day care centers and preschool programs, including Head Start.

ECPP-NHES:1995 included questions on children's participation in care or education provided by relatives, nonrelatives, Head Start programs, and center-based programs. It also collected information on the early school experiences of school-age children, home literacy activities, health and disability status, and parent and family characteristics.

The Early Childhood Education Survey (ECE-NHES:1991) included questions on participation in nonparental care or education; characteristics of programs and care arrangements; and early school experiences, including delayed kindergarten entry and retention in grade. In addition, parents were asked about activities that children engaged in with parents and other family members, inside and outside the home. Information on family, household, and child characteristics was also collected. Eligible respondents to this survey were the parents or guardians of the sampled 3- to 8-year-olds who were most knowledgeable about the children's education.

The School Readiness Survey of 2007 (SR-NHES:2007) collected information on early learning and readiness for entering school: specifically, participation in preschool or other types of center-based care and education, including Head Start; children's developmental accomplishments, including literacy and numeracy skills; educational activities with family members; plans for kindergarten enrollment; and the role of parents in preparing their child for kindergarten. The survey also collected data on the amount and type of television viewing by preschoolers.

SR-NHES:1993 included questions on the developmental characteristics of preschoolers; school adjustment, and teacher feedback to parents, for kindergartners and primary school students; center-based program participation; early school experiences; home activities with family members; and health status. Extensive information was collected on family and child background characteristics—including parents' language and education, income, receipt of public assistance, and household composition—to permit the identification of at-risk children. Eligible respondents to this survey were the parents or guardians of the sampled children (ages 3 through 7 in 2nd grade or below and children ages 8 and 9 in 1st or 2nd grade) who were most knowledgeable about the children's education.

Household Library Use. The Household and Library Use Survey (HHL-NHES:1996) was part of the 1996 NHES screener and consisted of a brief set of questions regarding public library use. Questions addressed the distance to the closest public library, household use of a public library in the past month and year, ways in which the public library was used, purposes for which the public library was used, and detailed household characteristics. Eligible respondents were those adults who completed the screener interview.

Parent and Family Involvement in Education. Surveys on this topic were conducted in 2007, 2003, 1999, and 1996.

The 2007 Parent and Family Involvement in Education Survey (PFI-NHES:2007) collected information on school choice, homeschooling, school characteristics (including school type, lowest and highest grades at the school, school religious affiliation, and whether the school was a magnet or charter school), student experiences in school, teacher feedback on the child's school performance and behavior, family involvement in school, family help with homework, family involvement in activities outside of school, factors affecting family involvement, and community support.

PFI-NHES:2003 focused on children and youth in kindergarten through grade 12 and addressed school experiences, family participation in schools, school practices to involve and support families, family involvement in schoolwork, and family involvement outside of school. Homeschooling parents were asked about their reasons for choosing, and resources for implementing, homeschooling. The involvement of nonresidential parents was also addressed, when applicable. In addition, information was collected on the child's or youth's health and disability status and on child and parent demographic characteristics. A total of 12,430 interviews were completed with parents of eligible children. The survey provided current national, cross-sectional estimates for the population of interest and provided the ability to examine change over time.

In 1999, the Parent Survey (Parent-NHES:1999) had six sets of questions, appropriate for six subgroups of children: children age 2 and younger, children ages 3 through 6 years and not yet in kindergarten, children in kindergarten through the 5th grade, youth in the 6th through 8th grades, youth in the 9th through 12th grades,

and children from age 5 through 12th grade who were receiving homeschooling. The survey included questions on the following topics (although not all were covered for all populations): demographic characteristics, current school- or center-based program enrollment status, center-based program participation before school entry, homeschooling, school characteristics, school readiness skills, participation in early childhood care and programs, training and support for families of preschoolers, parents' satisfaction with children's schools, children's academic performance and behavior, family involvement with children's schools and school practices to involve families, before- and after-school programs and nonparental care, parents' expectations about children's college plans and costs, family involvement in educational activities outside of school, child health and disability, parent/guardian characteristics, and household characteristics. The Parent Survey was administered to the parent or guardian most knowledgeable about the education of each sampled child from birth through 12th grade.

In 1996, the survey on parent involvement was combined with one on civic involvement, forming PFI/CI-NHES:1996. It included questions on the schools of the sampled children, communication with teachers or other school personnel, school practices to involve parents, children's homework and behavior, and learning activities with children outside of school with their families. Information was also collected on students' experiences in school, children's personal and demographic characteristics, household characteristics, and children's health and disability status. Eligible respondents were the parents or guardians of the sampled children ages 3 through 20 and in 12th grade or below who were the most knowledgeable about their education.

School Safety and Discipline. The 1993 School Safety and Discipline Survey (SS&D-NHES:1993) included questions on the school learning environment, discipline policy, safety at school, victimization, the availability and use of alcohol and drugs, and alcohol and drug education. The survey also included questions on peer norms for behavior in school and substance use. Extensive family and household background information and data about the characteristics of the school that the child attended were collected. Eligible respondents were the parents or guardians of the sampled children in grades 3 through 12 and youth in grades 6 through 12 who were most knowledgeable about the child's education.

Periodicity
NHES has been conducted in the spring of 1991, 1993, 1995, 1996, 1999, 2001, 2003, 2005, and 2007. NHES is currently undergoing a major redesign to address response rate and potential coverage issues. The next data collection is anticipated for 2012.

2. USES OF DATA

NHES provides descriptive data on the educational activities of the U.S. population and offers policymakers, researchers, and educators a variety of statistics on the condition of education in the United States. Each NHES survey collects specific data based on a set of research questions that guide the development of the questionnaire. As described above, the main subject areas for the NHES program are:

- Adult education and lifelong learning;
- Before- and after-school programs and activities;
- Early childhood education and school readiness; and
- Parent and family involvement in education; and

Analysts should review the instrument for each survey to identify areas of particular interest to them.

3. KEY CONCEPTS

See the survey documentation for definitions specific to any one NHES survey.

Household Members. Individuals who think of the sampled household as their primary place of residence, including persons who usually stay in the household but are temporarily away on business or vacation; in a hospital; or living at school in a dormitory, fraternity, or sorority.

4. SURVEY DESIGN

Target Population
Noninstitutionalized, civilian members of households in the 50 states and the District of Columbia. Because the topical surveys change from one NHES to the next,

the specific age or grade criteria for the target populations also change. In general, there are three educational populations of interest: (1) younger children from birth through 5th grade; (2) older children (i.e., youth) in the 6th through 12th grades; and (3) adults not enrolled in 12th grade or below. The respondent is usually the parent or guardian of the child who is most knowledgeable about the education or care of the sampled child, the sampled youth, or the sampled adult.

Sample Design

The NHES samples are selected using random-digit-dialing (RDD) methods. Telephone numbers are randomly sampled, and a screener is administered to sampled households. About 45,000 to 64,000 households are screened for each administration. Individuals within households who meet predetermined criteria are then sampled for more detailed or extended interviews.

Sampling Households. Two general sampling approaches have been taken: list assisted and a modified Mitofsky-Waksberg method. The list-assisted method has been used since the 1995 administration.

The sampling frame for NHES:2007, NHES:2005, and NHES:2003 was all telephone numbers in 100-banks (i.e., sets of numbers with the same first 8 digits of the 10-digit telephone number) with one or more listed residential telephone numbers as of the third quarter of 2006, September 2004, and September 2002, respectively. A stratified two-phase list-assisted sample was used in order to support design goals for national-level and subdomain statistics for the NHES surveys.

NHES:2007. In the first phase of sampling, a sample of 476,170 telephone numbers was drawn, with telephone numbers in areas with high percentages of Black or Hispanic residents sampled at higher rates than those in areas with low percentages. The sampling frame contains estimates of race/ethnicity distributions from the 2000 census, which are used to identify high concentrations of Black or Hispanic telephone exchanges. The sampling rate in the high- Black or Hispanic concentration stratum was nearly twice that in the low-Black or Hispanic stratum.

In the second phase, within each race/ethnicity stratum, the sampled telephone numbers were stratified as mailable or nonmailable according to whether a mailing address was able to be matched to the telephone number. Mailable status was used because it has been found to improve the efficiency of the sample by facilitating the oversampling of mailable numbers (which are more likely to be residential). Within each of the four strata defined by the combinations of Black or Hispanic concentration and mailable status, telephone numbers were subsampled at different rates in order to attain the final phase 2 allocation. The phase 1 sample sizes were determined by calculating the minimum number of telephone numbers expected to be needed from each race/ethnicity stratum in order to attain the desired phase 2 sample sizes in the race/ethnicity-by-mailable strata, based on mailable distributions within each race/ethnicity stratum computed from NHES:2005. The screener unit response rate in 2007 was 52.8 percent

NHES:2005. In the first phase of sampling, a sample of 350,000 telephone numbers was drawn, with telephone numbers in areas with high percentages of Black and Hispanic residents sampled at higher rates than those in areas with low percentages. The sampling frame contained the Census 2000 counts of persons in the area by race and ethnicity. Race and ethnicity information was obtained for zip codes served by the telephone exchange and then aggregated. A 100-bank was classified in the high-Black or Hispanic concentration stratum if its population was either at least 20 percent Black or at least 20 percent Hispanic. The banks that did not meet this requirement were classified in the low-Black or Hispanic concentration stratum. The sampling rate in the high-Black or Hispanic concentration stratum was nearly twice that in the low-Black or Hispanic concentration stratum. While telephone exchanges do not correspond exactly to census tracts or blocks, this approach is still effective at increasing the sample yield for Blacks, Hispanics, and Asians.

In the second phase, within each Black or Hispanic stratum, the sampled telephone numbers were classified as mailable or nonmailable according to whether they could be matched to a mailing address in the white pages of the telephone directory or in another database. Within each of the four strata defined by the combinations of Black or Hispanic concentration and mailable status, telephone numbers were subsampled at different rates. In the low-Black or Hispanic stratum, telephone numbers in the mailable substratum were sampled at a rate about 72 percent higher than numbers in the nonmailable substratum; in the high-Black or Hispanic stratum, telephone numbers in the mailable substratum were sampled at a rate about twice as high as that used for numbers in the nonmailable substratum.

In this manner, a sample of 207,000 telephone numbers was initially selected for NHES:2005. The remaining 143,000 telephone numbers from the first phase sample of 350,000 were held in reserve. Assuming that 49

percent of the sampled telephone numbers would belong to households and assuming a screener unit response rate of 65 percent, it was expected that about 59,380 screening interviews would be completed. For example, 25,260 screeners were expected to be completed in stratum 1 (mailable, high-Black or Hispanic concentration). This was calculated by taking the final NHES:2005 phase 2 allocation to stratum 1 (51,490 telephone numbers) and multiplying it by the expected residency rate (84 percent) to get the approximate number of residential telephone numbers (43,250). For the 60 percent of residential numbers that were randomly designated to receive the standard protocol, a 69 percent expected response rate was used to estimate the expected number of completed screeners; for the remaining 40 percent, a 43 percent initial cooperation rate was used to estimate the expected number of completed screeners. These calculations resulted in a total of 25,260 expected completed screeners for stratum 1. However, after the release of the initial sample of 207,000 telephone numbers, it was determined that the residency rates in the mailable strata were lower than expected. Thus, an additional 34,000 telephone numbers, subsampled from the 143,000 numbers in the reserve sample at the same rates used for the original sample, were released. The total number of telephone numbers released for the study was 241,000, including the 34,000 reserve telephone numbers. The screener unit response rate was 67 percent, and the number of households with completed screening interviews was 58,140.

NHES:2003. In the first phase of sampling, a sample of 144,300 telephone numbers was drawn, with telephone numbers in areas with high percentages of Black and Hispanic residents sampled at higher rates than those in areas with low percentages of Black and Hispanic residents. The sampling frame used in the study contained the Census 2000 counts of persons in the area by race and ethnicity. A 100-bank was classified in the high-Black or Hispanic concentration stratum if its population was either at least 20 percent Black or at least 20 percent Hispanic. The banks that did not meet this requirement were classified in the low-Black or Hispanic concentration stratum. The sampling rate in the high-Black or Hispanic concentration stratum was nearly twice that in the low-Black or Hispanic stratum.

In the second phase, within each Black or Hispanic stratum, the sampled telephone numbers were classified as mailable or nonmailable according to whether they could be matched to a mailing address in the white pages of the telephone directory or in another database. Within each of the four strata defined by the combinations of Black or Hispanic concentration and mailable status, telephone numbers were subsampled at different rates. In the low-Black or Hispanic stratum, telephone numbers in the mailable substratum were sampled at a rate about 47 percent higher than numbers in the nonmailable substratum; in the high-Black or Hispanic stratum, telephone numbers in the mailable substratum were sampled at a rate about 63 percent higher than numbers in the nonmailable substratum.

In this manner, a sample of 109,800 telephone numbers was selected for NHES:2003. (The remaining 34,500 telephone numbers from the first-phase sample of 144,300 were held in reserve. The reserve sample was not used.) Assuming that 49 percent of the telephone numbers would belong to households and assuming a screener unit response rate of 69 percent, it was expected that about 37,000 screening interviews would be completed. However, the actual unweighted residency rate was 45 percent, and the screener unit response rate was 65 percent. Thus, the number of households with completed screening interviews was 32,050.

NHES:2001. In 2001, a two-phase list-assisted method was also used. In the first phase of sampling, telephone numbers were stratified according to the percent of Black or Hispanic residents in the exchange. Exchanges with at least 20 percent Blacks or at least 20 percent Hispanics were classified as high-Black or Hispanic, and all other exchanges were classified as low-Black or Hispanic. Telephone numbers in the high-Black or Hispanic stratum were sampled at a rate of about 1 in 810, and telephone numbers in the low-Black or Hispanic stratum were sampled at a rate of about 1 in 1,560. The first-phase sample of telephone numbers was processed using the Genesys ID-Plus process to identify nonworking and business numbers. As part of this process, the telephone numbers were matched to white pages listings, and the matches were flagged. Thus, for each telephone number in the first-phase sample, the listed status (i.e., whether or not it is listed in the white pages) is known. Within each race/ethnicity stratum, the telephone numbers in the first-phase sample were stratified according to the white pages listed status; the overall number of telephone numbers selected in phase 1 was 206,180.

In the second phase, telephone numbers within each of the four strata defined by the combinations of Black or Hispanic concentration and listed status were subsampled at different rates: 0.71 for the high-Black or Hispanic listed stratum; 0.95 for the high-Black or Hispanic unlisted stratum; 0.73 for the low-Black or Hispanic listed stratum; and 0.94 for the low-Black or Hispanic unlisted stratum. The total number of telephone numbers selected in phase 2 was 179,210.

In the 1995, 1996, and 1999 administrations of NHES, a list-assisted method was used. This approach involves selecting a simple random sample of telephone numbers from all telephone numbers in 100-banks that have at least one telephone number listed in the white pages (called the listed stratum). Telephone numbers in 100-banks with no listed telephone numbers (called the zero-listed stratum) are not sampled. Because the list-assisted approach is an unclustered design, it results in estimates with lower variances than the clustered alternative methods. However, this method also incurs a small amount of coverage bias because households in the zero-listed stratum have no chance of being included in the sample. (See "Coverage error," in section 5 below, for a discussion of coverage bias. See Casady and Lepkowski [1993] for a further description of the list-assisted method.)

For the surveys fielded in 1996, the goal of making estimates at the state level for characteristics of household members and for household library use also determined the number of telephone numbers selected. A target of 500 screened households per state was set. A sample of 500 households is large enough that, if 30 percent of the households in a state have a given characteristic, differences of 6 percent can be detected. Due to nonresponse at the screener level and lower residency rates than expected, 500 screeners were not completed in some states. The lower number of responses limits the ability to make estimates for some subgroups within states. Analysts should examine the standard errors for subgroups of interest to evaluate the precision of within-state estimates.

The NHES surveys fielded in 1991 and 1993 used a modified version of the Mitofsky-Waksberg method of RDD, in which a fixed number of telephone numbers is sampled from 100-banks. (See Brick and Waksberg [1991] for a further description of the modified Mitofsky-Waksberg method used in NHES.)

Oversampling households for Blacks and Hispanics. One of the goals of the NHES program is to produce reliable estimates for subdomains defined by race and ethnicity. In a 64,000-household design in which every household has the same probability of being included, the number of completed interviews would not be large enough to produce reliable estimates of many characteristics of Black and Hispanic youth. Therefore, in each NHES administration, telephone numbers in areas with high concentrations of Blacks and Hispanics are oversampled.[1]

A computer file containing census characteristics for telephone exchanges is used to stratify telephone exchanges into low- and high-Black or Hispanic concentration strata. Any telephone exchange not found in the file is assigned to the low-Black or Hispanic concentration stratum. High-Black or Hispanic concentration exchanges are defined as those having at least 20 percent Black or 20 percent Hispanic persons living in the area.[2] The telephone exchanges in the two strata are identified, and a systematic sample is drawn in each stratum. The sampling fraction used in the high-Black or Hispanic concentration stratum is two times the fraction used in the low-Black or Hispanic concentration stratum. Oversampling by the characteristics of the telephone exchange has two effects. First, the oversampling increases the sample sizes for Blacks and Hispanics because they are more heavily concentrated in the exchanges that are oversampled. Second, the sampling errors for estimates of these groups are reduced due to the increased sample sizes. On the other hand, not all race/ethnicity groups are found in the oversampled exchanges. Thus, differential sampling rates are applied to persons depending on their exchanges. Using differential rates increases the sampling errors of the estimates, partially offsetting the benefit of the larger Black and Hispanic sample. However, the net result is an increase in precision of estimates for Blacks and Hispanics. The technical report *Effectiveness of Oversampling Blacks and Hispanics in the NHES Field Test* (Mohadjer 1992) indicates that oversampling is successful in reducing the variances for estimates of characteristics of Blacks and Hispanics by approximately 20 to 30 percent over a range of statistics examined. The decreases in precision for estimates of the groups that are not oversampled and for estimates of totals are modest, ranging from about 5 to 15 percent.

Approaches to household enumeration. The approach to screening households has also changed over the course of the NHES program. Changes have been made in the methods of enumerating members of households that are contacted and the amount of information collected in the screener about the household and its members. In 1991, a split-enumeration design was used; all households were

[1] In 1993, areas with high percentages of Asians/Pacific Islanders were also sampled at a higher rate; this was discontinued in later administrations because the new vendor for numbers used in the list-assisted sampling did not have this information available. NHES considered reintroducing an Asian/Pacific Islander oversampling strategy in 2001. However, it was determined that more precision in other racial/ethnic groups would have been lost than was warranted, given the amount of extra precision that would have been gained for Asians/Pacific Islanders.

[2] For the 1993 NHES, high Asian/Pacific Islander concentration exchanges were defined as those having at least 20 percent Asian/Pacific Islander persons living in the area.

screened for ECE-NHES:1991, and a subset of households was screened for AE-NHES:1991. In 1993, when SR-NHES:1993 and SS&D-NHES:1993 were fielded, households were enumerated only when there were household members age 20 or younger. The only information collected in both 1991 and 1993 was the first name, age, and sex of household members. In both 1995 and 1996, all screened households were fully enumerated. The 1995 administration included a test of an expanded screener that was used in 1996, but dropped from later NHES administrations. The 1996 screener collected educational and demographic information on household members and included a brief topical survey. The 1999 screener again collected first name, age, and sex of household members, but not all households were fully enumerated; thus, if the screener respondent said there were no children in the household and the household had not been preselected for an adult education interview, the screener information was not collected.

Sampling within households. The within-household sample designs for the NHES collections are determined by the specific goals of the surveys administered and by the combination of surveys administered in a specific year. Brief summaries of the within-household sampling for the various NHES administrations are given below, by year.

2007 NHES surveys—SR-NHES:2007, and PFI-NHES:2007. Originally, an Adult Education for Work Related Reasons (AEWR) module was planned. The sampling scheme took this survey into account.The sampling scheme for within-household sampling was designed to satisfy the sample requirements discussed earlier, while keeping respondent burden to a minimum. To carry out this sampling scheme, several flags and/or random numbers were set prior to screening (i.e., at the time the sample of telephone numbers was drawn). The first specified whether the adult sampling algorithm was to be run for a particular household (in order to determine whether an adult was to be selected). Each telephone number received one of three possible designations: household was designated for the adult sampling algorithm to be run; household was designated for the adult sampling algorithm to be run only if there were no eligible children in the household; or household was not designated for the adult sampling algorithm to be run. The expected number of completed screeners for stratum 1 was calculated in the following manner: First, the final NHES:2007 phase 2 allocation to stratum 1 (74,480 telephone numbers) was multiplied by the expected residency rate for cases in this stratum (73 percent) to get the expected number of residential telephone numbers in stratum 1 (54,370). Next, for the 60 percent of those residential numbers that were randomly designated to receive the standard protocol, a 63 percent expected response rate was applied to the expected number of residential telephone numbers; for the remaining 40 percent, a 39 percent initial cooperation rate was applied. These calculations resulted in a total of 29,190 expected completed screeners for stratum 1.

Once the enumeration of the appropriate household members was completed in the screener, the sampling of household members for the extended interviews was done by computer. The PFI interviews were conducted with the parents or guardians of sampled children and youth in kindergarten through 12th grade with a maximum age of 20. Following the enumeration of children, if the household had at least one preschooler, then exactly one was randomly sampled for the SR survey. If the household had at least one child ages 3 through 20 enrolled in kindergarten through 12th grade, then exactly one was randomly sampled for the PFI survey. For each survey, pre-assigned random numbers were used to sample from among all eligible children in the household. In households in which an adult was sampled, adult education participants had twice the probability of selection of nonparticipants.

2005 NHES surveys—ECPP-NHES:2005, ASPA-NHES:2005, and AE-NHES:2005. To limit respondent burden, a within-household sampling scheme was developed to control the number of persons sampled for extended interviews in each household. In all households with children age 15 or younger, children were enumerated. To determine whether adults would be enumerated, the sample of telephone numbers was randomly divided into three groups. The first group (80,850 telephone numbers, or approximately one-third of the sample) was designated for adult enumeration. The second group (40,070 telephone numbers, or about one-sixth of the sample) was designated for adult enumeration only if there were no eligible children in the household. The third group (120,080 telephone numbers, or about one-half of the sample) was designated for no adult enumeration.

Once the enumeration of the appropriate household members was completed in the screener, the sampling of household members for the extended interviews was done by computer. The ECPP and ASPA interviews were conducted with the parents or guardians of sampled children from birth through age 15 who were in grade 8 or below. In households with one or more preschoolers (children age 3 through 6 and not yet in kindergarten), one child in this age/grade range was sampled. In households with

middle school students (6th through 8th grade), one child in this age/grade range was also sampled. The sampling of infants (newborn through age 2), elementary school children (kindergarten through 5th grade), and adults was conducted using an algorithm designed to attain the sampling rates required to meet the target sample sizes while minimizing the number of interviews per household. The within-household sample size was limited to three eligible children (if no adults were to be selected) or to two eligible children and one eligible adult. No more than one child from any given domain (i.e., infants, preschoolers, elementary students, middle school students) was sampled in any given household. This sampling algorithm was designed to limit the amount of time required to conduct interviews with parents in households with a large number of eligible children. If no children were selected and there were multiple adults with less than a high school diploma or the equivalent, up to two adults could be selected.

2003 NHES surveys—PFI-NHES:2003 and AEWR-NHES:2003. Sampling within households for NHES:2003 followed a similar methodology as in 2005. In all households with children and youth age 20 or younger, children and youth were enumerated. To determine whether adults would be enumerated, the sample of telephone numbers was randomly divided into three groups. The first group (63,620 telephone numbers, or approximately 44 percent of the sample) was designated for adult enumeration. The second group (63,730 telephone numbers, or about 44 percent of the sample) was designated for adult enumeration only if there were no eligible children or youth in the household. The third group (16,950 telephone numbers, or about 12 percent of the sample) was designated for no adult enumeration.

Once the enumeration of the appropriate household members was completed in the screener, the sampling of household members for the extended interviews was done by computer. The PFI interviews were conducted with the parents or guardians of the sampled children and youth in kindergarten through 12th grade (with a maximum age of 20). If there were one or two eligible children or youth, all were selected with certainty. In households with more than two eligible children or youth, two were selected with equal probability. The sampling of adults was conducted using an algorithm designed to attain the sampling rates required to meet the target sample sizes while minimizing the number of interviews per household. The within-household sample size was limited to two eligible children and one eligible adult. This sampling algorithm was designed to limit the amount of time required to conduct interviews with parents in households with a large number of eligible children.

2001 NHES surveys—AELL-NHES:2001, ASPA-NHES:2001, and ECPP-NHES:2001. A within-household sample scheme was developed to control the number of persons sampled for extended interviews in each household. The sample of telephone numbers was randomly divided into three groups. The first group (89,600 telephone numbers, or approximately 50 percent of the sample) was designated for adult enumeration. The second group (44,990 telephone numbers, or about 25 percent of the sample) was designated for adult enumeration only if there were no eligible children in the household. The third group (44,630 telephone numbers, or about 25 percent of the sample) was designated for no adult enumeration. Once the enumeration of the appropriate household members was completed in the screener, the sample of household members for the extended interviews was done by computer. The ECPP and ASPA interviews were conducted with the parents or guardians of sampled children from birth through age 15 who were in 8th grade or below. In households with one or more preschoolers (children age 3 through 6 and not yet in kindergarten), one child in this age/grade range was sampled. In households with middle school students (6th through 8th grades), one child in this age/grade range was also sampled. The sampling of infants (newborn through age 2), elementary school children (kindergarten through grade 5), and adults was conducted using an algorithm designed to attain the sample rates required to meet the target sample sizes while minimizing the number of interviews per household. The within-household sample size was limited to three eligible children (if no adults were to be selected) or to two eligible children and one eligible adult. No more than one child from any given domain (i.e., infants, preschoolers, elementary students, middle school students) was sampled in any given household. This sampling algorithm was designed to limit the amount of time required to conduct interviews with parents in households with a large number of eligible children.

1999 NHES surveys—AE-NHES:1999, Parent-NHES:1999, and Youth-NHES:1999. The overall screening sample was largely determined by the need to produce precise estimates of indicators for young children, particularly preschoolers. Since sample requirements were most stringent for preschoolers (children ages 3 through 6 not yet in kindergarten), it was decided to sample one preschooler in every household with preschoolers. Another goal was that no more than three persons per household be sampled, with a maximum of four extended interviews per

household. To accomplish this, several flags were set prior to screening. The first specified whether adults in the household were to be enumerated, as well as the conditions under which an adult was to be sampled. This flag was set such that households without eligible children or youth were sampled for an Adult Education Survey at approximately twice the rate of households with eligible children or youth (about 26 percent vs. 13 percent). Additionally, this flag enabled one- and two-adult households with no adult education participants to be further subsampled at a fixed, prespecified rate (25 percent for one-adult households and 75 percent for two-adult households). The second flag designated whether an infant was to be sampled, if the household had two other sampled children or youth. A third flag designated whether a younger child or an older child was to be sampled; if the household had children in both groups, only one was to be selected. In households in which an adult was to be sampled, each adult education participant was given a probability of selection 2.5 times as large as the probability of selection assigned to nonparticipants.

1996 NHES surveys—ACI-NHES:1996, HHL-NHES:1996, PFI/CI-NHES:1996, and YCI-NHES:1996. The number of interviews for which household members could be selected was limited by creating two separate samples—parent/youth and adult. A sample of 161,450 telephone numbers was selected and randomly divided into two groups. The first group (153,370 telephone numbers, or 95 percent of the sample) was allocated to the parent/youth sample. A screening interview was conducted in these households, and eligible children and youth were sampled, respectively, for PFI/CI-NHES:1996 or for both PFI/CI-NHES:1996 and YCI-NHES:1996. For PFI/CI-NHES:1996, if there were one or more children from age 3 through 5th grade (younger children), one child in this age range in the household was sampled for the survey. If the household included one or more children in 6th through 12th grades (older children), one child in this grade range in the household was sampled for the survey. If an older child was sampled as the subject of a PFI/CINHES:1996 interview, the child was also asked to complete YCI-NHES:1996. Because households may have had up to two parent PFI/CI interviews (one for a younger child and one for an older child), the maximum number of interviews per sampled household was three. The other group (8,070 telephone numbers, or 5 percent of the sample) contained those telephone numbers allocated to ACI-NHES:1996. For households in this group, a screening interview was conducted ACI-NHES:1996 was administered to one eligible adult.

1995 NHES surveys—AE-NHES:1995 and ECPP-NHES:1995. Interviews for ECPP-NHES:1995 were conducted with the parents or guardians who were the most knowledgeable about the education of the sampled children aged 0 to 10 and in the 3rd grade or below. The within-household sample size was limited to two eligible children. Children in kindergarten were sampled at 1.5 times the rate for other children to improve the precision of single-year estimates for kindergartners. Any adult aged 16 years or older not currently enrolled in secondary school was eligible for sampling for AE-NHES:1995. Sampled adults who said they were on active duty in the U.S. Armed Forces were classified as ineligible for the interview.

1993 NHES surveys—SR-NHES:1993 and SS&D-NHES:1993. For the 1993 NHES surveys, children within households were subsampled. For SR-NHES:1993, interviews were conducted with the parents or guardians who were the most knowledgeable about the education of the sampled children aged 3 through 7 (as well as children aged 8 or 9 who had not completed 2nd grade. If there were one or two eligible children in a household, all were sampled. If there were more than two eligible children in a household, two were randomly sampled. Any child enrolled in grades 3 through 12 and below the age of 21 was eligible for sampling for the SS&D-NHES:1993 interview with the parent. Sampling was limited to one child in 3rd through 5th grades and no more than two children in any household. No more than one youth was subsampled per household for the youth interview. If a child was enrolled in the 6th through 12th grades but did not live with a parent or guardian, he or she was considered an emancipated youth. A special emancipated youth interview was conducted, including some questions usually asked only of parents.

1991 NHES surveys—AE-NHES:1991 and ECE-NHES:1991. All 3- to 8-year-olds in sampled households were included in ECE-NHES:1991, as were 9-year-olds who had not completed 2nd grade. This ensured that nearly all children eligible for the extended interviews were identified, even if a rounding error was made in reporting the children's ages. The respondent for the interview was the parent or guardian of the sampled child reported to be the most knowledgeable about the child's care and education. Only a subset of households was screened for AE-NHES:1991. In the screened households, all adults identified as participating in adult education activities were sampled, half of the full-time degree-seeking students were sampled, and about 7 percent of the nonparticipants in adult education activities were sampled. After a few weeks of data collection, the number of sampled households screened for AE-

NHES:1991 was reduced because the required number of interviews had been completed; thus, additional households did not need to be contacted. Altogether, 18,460 households out of 60,300 completed screeners (31 percent) were sampled for AE-NHES:1991. In addition, the sampling rate for nonparticipants was increased from 7 to 12 percent.

Data Collection and Processing

NHES program surveys are conducted using computer-assisted telephone interviewing (CATI). Westat has been the contractor on all surveys to date.

Reference dates. Most data items refer to the time of data collection or to the interval of time between the data collection and September of the current school year. Other items are asked retrospectively for different time frames. For example, in the 1996 NHES surveys, respondents were asked about family involvement with children outside of school (e.g., reading with a child, visiting a library) in the past week and past month; civic involvement (reading about or watching national news) in the past week; political activities in the past 12 months; voting activities in the past 5 years; working for pay during the past week and the past 12 months; job hunting in the past 4 weeks; child's communications with the noncustodial parent in a typical month and in the past year; youth's discussion of future educational plans with parents in the past month; books read in the past 6 months; home visits by professionals during the past 12 months; and religious service participation in the past year. The adult education information is based on participation in the past 12 months.

Data collection. Data collection for the NHES surveys takes place over a 3- to 4-month period beginning in January of each survey year. The data are collected using CATI. The NHES screeners are completed with an adult household member in households selected using RDD techniques. (See "Sample Design" in section 4 above.)

Over a period of about 3 weeks just prior to data collection, more than 300 interviewers undergo intensive training in general interviewing techniques, use of the CATI system, and the conduct of the survey.

Most responses to survey items are coded at the time of the interview. Most of the items are close ended, meaning respondents are given a short list of response options. Interviewers simply record the response as a one- or two-digit code that is entered directly into the data file as the interview progresses. However, most close-ended items do have "other, specify" options that allow interviewers to record responses that do not fit the precoded response categories. The interviewer types in these open-ended responses as one or more sentences. "Other, specify" responses to close-ended items are rare.

A small number of items in some of the surveys are designed to be open ended. That is, precoded categories do not exist and interviewers type in verbatim responses from respondents. Once the survey is completed, data preparation staff and survey managers review these open-ended responses to determine how they can be coded into a limited set of response categories. Coding of additional open-ended items was required for the Adult Education Surveys administered in 1991 and 1995. These items were for adult education courses, major fields of study for college and vocational programs, industry, and occupation. A double-blind coding procedure was used, in which two coders independently assigned a code to the response. When the coding was discrepant, an "adjudication" coder reviewed the case and assigned an appropriate final code.

Editing. Intensive data editing is a feature of both the data collection and file preparation phases of the NHES collections. Range checks for allowable values and logic checks for consistency between items are included in the online CATI interview so that many unlikely values or inconsistent responses can be resolved while the interviewer is speaking with the respondent.

Postinterview editing is conducted throughout data collection and after data collection is completed. In addition to range and logic edits, the postinterview editing process includes checks for the structural integrity of the hierarchical CATI database and integrity edits for complex skip patterns. It also includes a review of comments provided by interviewers and problem sheets completed by interviewers. Following the resolution of any problems, data preparation staff review frequency distributions and cross-tabulations of the datasets in order to identify any remaining skip pattern inconsistencies. Editing is repeated following completion of imputation.

Estimation Methods

The NHES surveys use weighting to adjust for the fact that the sampling method used is not simple random sampling. It is also used to adjust for potential undercoverage bias and potential unit nonresponse bias. Imputation is performed to compensate for item nonresponse.

Weighting. The objective of the NHES surveys is to make inferences about the entire noninstitutionalized, U.S. civilian population and about subgroups of

interest. Although only telephone households are sampled, the estimates are adjusted to totals of persons living in both telephone and nontelephone households derived from the Current Population Survey (CPS) to achieve this goal. (CPS is an annual household survey conducted by the U.S. Bureau of the Census for the U.S. Bureau of Labor Statistics.) As a result, any undercoverage in CPS for special populations, such as the homeless, is also reflected in NHES estimates. The potential for bias due to sampling only telephone households has been examined for virtually all the population groups sampled in NHES. Generally, the bias in the estimates due to excluding nontelephone households is small. (See "Coverage error" in section 5 below for further discussion.) The weighting procedures across NHES surveys are very similar. Weighting consists of two stages: household-level weighting and person-level weighting, as described below.

Household weights. The household weights take into account all factors that might have resulted in adjustments due to the telephone numbers being sampled at different rates. Two factors common to all NHES years are (1) the adjustment to account for the differential sampling rates by Black or Hispanic concentration; and (2) the adjustment to account for households that have more than one telephone number and, hence, have a greater chance of being sampled. In 1991 and 1993, an adjustment was also made to account for the modified Mitofsky-Waksberg method of RDD sampling. (See "Sample Design" in section 4 above.) The 1996 NHES included an adjustment for the oversampling in 18 states to bring the minimum expected number of completed screeners up to 500.

In NHES 2007, the primary purpose of the screener was to provide the information required to assess the eligibility of household members for an extended interview. Household-level information that is of analytic interest was also collected during the extended interview. Since no data intended for analyses were collected at the household level only, household-level weights were calculated solely for use as a basis for computing person-level weights for the analysis of the extended interview data. The household-level weight was the product of five factors:

1. The weight associated with the differential sampling of telephone numbers based on the Black or Hispanic stratum of the exchange and the mailable status of the telephone number;

2. An adjustment for subsampling of cases for nonresponse follow-up;

3. An adjustment for the subsampling of screener nonresponse cases;

4. An adjustment for the number of telephone numbers in a household; and

5. A poststratification adjustment to compensate for the fact that only landline telephone households were eligible for the NHES:2007 surveys.

The calculation of the household weight, taking into account these five factors, is discussed below.

The first step was to assign the weight associated with the differential sampling of telephone numbers based on the Black or Hispanic concentration stratum of the exchange and the mailable status of the telephone number. The RDD sampling method used in NHES:2007 was a list-assisted method (the same basic method as was used in NHES:1995, NHES:1996, NHES:1999, NHES:2001, NHES:2003, and NHES:2005). In NHES:2007, as in NHES:2001, NHES:2003, and NHES:2005, a two-phase approach was used. In the first phase, a single-stage sample of telephone numbers was selected from strata defined by the Black or Hispanic concentration status of the exchange. Telephone numbers in high-Black or Hispanic exchanges were sampled at a rate approximately twice that of those in low-Black or Hispanic exchanges. An attempt was made to match each telephone number selected in the first phase to an address listing. In the second phase, telephone numbers were subsampled differentially within each Black or Hispanic concentration stratum based on the mailable status (i.e., whether a mailing address was obtained for the telephone number).

The second step in creating the household weight was to adjust for the subsampling of screener nonresponse cases.

The third weighting factor adjusted for households that did not respond to the NHES:2007 screener.

The fourth step in adjusting the household weight was to adjust for the number of telephone numbers in a household. A weighting factor of one was assigned to households reporting one telephone number in the household. An adjustment factor of one-half was assigned to households with exactly two residential telephone numbers, and an adjustment factor of one-third was assigned to households with three or more residential telephone numbers. Technically, if the other telephone numbers of households with multiple residential telephone numbers are in the zero-listed

stratum, the household should get a weight adjustment of one. However, looking up the other numbers to determine whether each is in the zero-listed stratum is impractical, and the percentage of such numbers in the zero-listed stratum is small.

The final step in computing the household weight was to account for household-level undercoverage due to sampling only landline telephone households. Poststratification was used to accomplish this task.

Person weights. The second stage of weighting forms person weights for each extended interview. The household-level weight was used to compute the base weight for each of the person-level (SR and PFI interview) weights in NHES:2007. The person-level weight for sampled person k in household j, PW_{jk}, is the product of the household weight and four weight adjustment factors:

1. The weight associated with sampling the person's domain in the given household;

2. The weight associated with sampling the person from among all eligible persons in the given domain in the household;

3. The weight associated with extended interview (SR or PFI) unit nonresponse; and

4. An adjustment associated with raking the person-level weights to Census Bureau estimates of the number of persons in the target population.

The development of the person-level weights, taking into account these four factors, is discussed below.

The first step was to account for the probability of sampling the person's domain in the given household. For both the SR and PFI interviews, if there was an eligible child in the household, then at least one child was selected; however, only one child was sampled for each survey in households with eligible children. Thus, the factor for sampling in both the SR and PFI domain was always equal to 1.

The second adjustment accounted for the probability of sampling the person from among all eligible persons in the given domain in the household. For each sampled person, the unadjusted person-level weight can be written as the product of the household-level weight and the adjustments for within-household sampling.
The third step was to adjust for persons who did not respond to the extended interview (i.e., the most knowledgeable parents or guardians in the case of the SR and PFI interviews). Each extended interview case was classified as either a respondent or a nonrespondent, depending on whether or not the extended interview was completed for the sampled person. The unadjusted person-level weights of the nonrespondents were distributed to the unadjusted person-level weights of the respondents within a nonresponse adjustment cell. For the SR and PFI Surveys, the nonresponse adjustment cells were created using combinations of home tenure (owned or rented), the four census regions, and age/grade combinations: unenrolled children age 3 through 6, preschoolers, kindergarteners, and children enrolled in each single grade for grade 1 through grade 12. (Enrolled children with no grade equivalent were included in the cell containing the modal grade for their age; that is, they were assigned to the grade in which most children their age are enrolled.) For PFI, whether the child attended regular school or was home schooled was also used. These variables were used because they are available for all sampled children (both respondents and nonrespondents) and are associated with SR/PFI interview response propensity.

The final stage of person-level weighting involved raking the nonresponse-adjusted person-level weights to national control totals. The raking procedure is carried out in a sequence of adjustments: first, the base weights are adjusted to one marginal distribution (or dimension) and then the second marginal distribution, and so on. One sequence of adjustments to the marginal distributions is known as a cycle or iteration. The procedure is repeated until convergence of weighted totals to all sets of marginal distributions is achieved. This additional raking adjustment, following the household-level poststratification adjustment, is required because the extended interviews involve new eligibility criteria and a new level of sampling. That is, although the household-level poststratification adjustment aligned the weighted totals of the household weights with the household-level control totals, the raking of the person-level weights is required in order to align the person-level weights with the person-level control totals and adjust for differential coverage rates at the person level.

The raking procedure for the SR and PFI weights involved raking the nonresponse-adjusted person-level weights to national totals obtained using the percentage distributions from the October 2005 CPS and the total number of children from the March 2006 CPS.

Imputation. Item response rates for most data items collected in NHES surveys are very high. Nevertheless, virtually all items with missing data (including "don't know" and "refused" responses) are imputed in NHES

surveys. In the two NHES surveys administered in 1991, only variables that were used for the development of weights or derived variables were fully imputed. Text responses (for example, in Youth-NHES:1999, type of service activity, or, in AE-NHES:1999, name of company) were not imputed in any year. Occasionally, “don’t know” and “refused” responses are of analytic interest, so they are not imputed. For example, in the Youth-NHES:1999 survey, “don’t know” and “refused to answer” responses to the knowledge about government items were not imputed.

In NHES:2007, for the SR and PFI Surveys, the median item response rates were 99.28 percent and 99.04 percent, respectively, and the median total response rates (the product of the item response rate and overall unit response rate) were 40.41 percent and 38.72 percent, respectively. Numeric and categorical data items with missing data in the file were imputed. (In general, character string variables, such as countries of origin, languages, or “other/specify” responses, were not imputed. School characteristics merged to the PFI data file from the NCES Common Core of Data [CCD] and Private School Universe Survey [PSS] files also were not imputed.)

Imputations are done in the NHES program for three reasons. First, complete responses are needed for the variables used in developing the sampling weights. Second, data users compute estimates employing a variety of methods, and complete responses should aid their analysis. Third, imputation may reduce bias due to item nonresponse, by obtaining imputed values from donors that are similar to the recipients. The procedures for imputing missing data are discussed below.

A standard (random within-class) hot-deck procedure has been used to impute missing responses in every NHES data collection. The methodology used for imputation in NHES:2007 was very similar to that used in previous NHES survey administrations. (The NHES:2007 procedures were based on those used in NHES:1996, NHES:1999, NHES:2001, NHES:2003, and NHES:2005.) In the hot-deck approach, the entire file is sorted into cells defined by characteristics of the respondents. The variables used in the sorting are general descriptors of the interview and include any variables involved in the skip pattern for the items. All of the observations are sorted into cells defined by the responses to the sort variables, and then divided into two classes within the cell depending on whether or not the item being imputed is missing. For an observation with a missing value, a value from a randomly selected donor (with the item completed) is used to replace the missing value. After the imputation is completed, edit programs are run to ensure that the imputed responses do not violate edit rules.

For some items, the missing values are imputed manually rather than using the hot-deck procedure. In NHES:2007, manual imputation was done (1) to impute certain person-level demographic characteristics; (2) to impute whether a child is homeschooled, whether the child attends regular school for some classes, and the number of hours the child attends regular school; (3) to correct for a small number of inconsistent imputed values; and (4) to impute for a few cases when no donors with matching sort variable values could be found.

Some person-level characteristics from the screener as well as from several sections of the SR and PFI interviews (age confirmation, household relationships, and child and parent language) were imputed manually because they typically involve complex relationships and/or constraints that would have required extensive programming in order to impute using a hot-deck procedure. The same is true of the items indicating whether a child is homeschooled, whether the child attends regular school for some classes, and the number of hours the child attends regular school. Furthermore, the reasonableness of imputed values for these person-level characteristics can often be assessed by examining the values of these variables for other members of the household. The use of the manual imputation approach in this situation permits the review of the characteristics of household members when imputing the missing values for the person-level variables.

After values have been imputed for all observations with missing values, the distribution of the item prior to imputation (i.e., the respondent’s distribution) is compared to the post-imputation distribution of the imputed values alone and of the imputed values together with the observed values. This comparison is an important step in assessing the potential impact of item nonresponse bias and ensuring that the imputation procedure reduces this bias, particularly for items with relatively low response rates (less than 90 percent).

For each data item for which any values are imputed, an imputation flag variable is created so that users can identify imputed values. Users can employ the imputation flag to delete the imputed values, use alternative imputation procedures, or account for the imputation in computation of the reliability of the estimates produced from the dataset.

Recent Changes

A two-phase sample design was used in the NHES surveys administered in 2001, and the NHES program adopted a new procedure for replication variance estimation for two-phase samples.

Future Plans

NHES is currently undergoing a major redesign to address falling response rates and potential coverage issues in the landline list assisted RDD design. The proposed new design utilizes an Address Based Sample (ABS) and will primarily collect data using a self administered paper questionnaire that will be mailed to sampled households. The first full scale data collection under the new design in anticipated to take place January 2012.

5. DATA QUALITY AND COMPARABILITY

In addition to the data quality activities inherent in the NHES design and survey procedures, activities designed specifically to assess data quality are undertaken for each collection. Reinterviews and analysis of telephone coverage bias are two activities conducted during many survey administrations. Other data quality activities address specific concerns related to a topical survey. Issues of data quality and comparability are discussed below.

Sampling Error

The two major methods of producing approximate standard errors for complex samples are replication methods and Taylor Series approximations. Special software is available for both methods, and the NHES data support either type of analysis. (Further information on the use of replication and Taylor Series methods is provided in *A Guide to Using Data From the National Household Education Survey* [Collins and Chandler 1997].)

Since the 2001 NHES surveys used a two-phase sample design, a new procedure for replication variance estimation was used thereafter. The replicate base weights under two-phase sampling are calculated using a two-step procedure. First, the initial replicate base weights of the first-phase units are calculated using the standard jackknife procedure. In the second step, the final replicate base weights for the second-phase sample are computed by redistributing the initial replicate weights of first-phase units not selected in the second phase to the initial replicate weights of the second-phase units within the same second-phase stratum.

Note that the sum of the final replicate base weights of the second-phase units is the same as the sum of the initial replicate base weights of the first-phase units within the same second-phase stratum. The procedure involves only the calculation of the telephone number-level replicate base weights. All full-sample weighting and all subsequent adjustments to the replicate weights are done using the same methodology used for a single-phase sample.

The replication method used in the NHES surveys for single-phase samples involves splitting the entire sample into a set of groups, or replicates, based on the actual sample design of the survey. The survey estimates can then be estimated for each of the replicates by creating replicate weights that mimic the actual sample design and estimation procedures used in the full sample. The variation in the estimates computed from the replicate weights can then be used to estimate the sampling errors of the estimates from the full sample. The procedures used to develop the full weights are used to produce each replicate weight. Replicate weights have been included in all of the NHES data files to make this application relatively simple. Various software packages, such as WesVar and SUDAAN, can properly apply replicate weights.

Nonsampling Error

Sample estimates also are subject to bias from nonsampling errors; however, it is more difficult to measure the magnitude of these errors. They can arise for a variety of reasons: nonresponse; undercoverage; differences in respondents' interpretations of the meaning of questions; memory effects; misrecording of responses; incorrect editing, coding, and data entry; time effects; or errors in data processing.

Coverage error. Every household survey is subject to some undercoverage bias—the result of some members of the target population being either deliberately or inadvertently missed in the survey. Telephone surveys like those in the NHES program are subject to an additional source of bias because not all households in the United States have telephones. Even more problematic is the fact that the percentage of households without telephones varies from one subgroup of the population to another. Differential rates among population subgroups, such as those defined by region, age, race/ethnicity, and household composition, are of concern to telephone survey methodologists because they can introduce bias in the estimates. Coverage bias in the telephone survey is probably due to the prevalence of nontelephone households (nontelephone households include cellular phone-only households, in addition to households with

no telephone service) and the differences between such households and those with telephones.

Based on recent findings (Blumberg and Luke, 2010) 24.5 percent of households had only a wireless telephone in 2009. Tucker et al. (2002) and Blumberg et al. (2006) examined differences in characteristics among persons and households having no telephone service, cellular service only, and landline service (including both landline only, and landline and cellular). Although there are differences in landline coverage (e.g., young adults, adults in 1-person households, renters, and Hispanics have lower landline coverage rates than other groups), raking is used in NHES to statistically adjust for and reduce undercoverage bias.

Special analyses of the bias associated with telephone coverage and its potential impact on estimates from the NHES surveys are conducted for each cycle of the survey. CPS data are used to evaluate the differences between estimates for telephone households and estimates for the entire population. The results of these analyses show that, for most estimates, the bias due to sampling only telephone households is small. However, for subgroups with characteristics highly correlated with not having a telephone (e.g., the poor, high school dropouts), coverage bias may be large. Recent studies suggest that between 5-20 percent of the population may be missed by using list assisted RDD methods (Boyle et al. and Fahimi et al.). Raking adjustments can reduce such coverage bias, though no adjustments have been found to adequately reduce the amount of bias across all measures that might be affected by coverage issues. Additionally, as the coverage bias increases, it becomes more difficult for raking to adequately adjust (See, for example, Montaquilla, Brick, and Brock [1997].)

Additional undercoverage results when some telephone households are excluded from the sampling frame. This was a disadvantage of the list-assisted method of RDD sampling used in earlier administrations of NHES surveys. (See section 4. “Survey Design,”above.) Households in the zero-listed stratum had no chance of being included in the sample. Empirical findings that address questions of coverage bias show that the percentage of telephone numbers in the zero-listed stratum that are residential is very small (about 1.4 percent) and that about 3 to 4 percent of all telephone households are in the zero-listed stratum. The findings also show that the bias resulting from excluding the zero-listed stratum is generally small. (See Brick et al. [1995].)

Nonresponse error. Nonresponse in NHES surveys is handled in ways designed to minimize the impact on data quality—through weighting adjustments for unit nonresponse and through imputation for item nonresponse.

Unit nonresponse. Household members are identified for extended interviews in a two-stage process. First, screener interviews are conducted to enumerate and sample households for the extended interviews. The failure to complete the first-stage screener means that it is not possible to enumerate and interview members of the household. The completion rate for the first stage is the percentage of screeners completed by households. The completion rate for the second stage is the percentage of sampled and eligible persons with completed interviews. The survey response rate is the product of the first- and second-stage completion rates (screener completion rate x interview completion rate = survey response rate). All of the rates are weighted by the inverse of the units’ probability of selection (see table 19).

Item nonresponse. For most of the items collected in the NHES surveys, the item response rate is high. The median item response rate for items with any missing values for the surveys administered in 1995, 1996, and 1999 ranged from 98.4 to 99.5 percent, except for HHL-NHES:1996, where the median response rates for imputed items was 95.0 percent for household-level characteristics and 99.5 percent for person-level characteristics. For SR-NHES:1993, three items had response rates of less than 95 percent; for SS&D-NHES:1993, there were two such items. None of the ECE-NHES:1991 items had response rates of less than 94 percent, while most of the AE-NHES:1991 items had response rates of more than 99 percent; however, there was one item from the 1991 screen that had a response rate of 92 percent. For SR-NHES:2007 and PFI-NHES:2007, the median item response rates were 99.28 percent and 99.04 percent, respectively, and the median total response rates (the product of the item response rates and overall unit response rates) were 40.41 percent and 38.72 percent, respectively.

Measurement error. In order to assess item reliability and inform future NHES surveys, most administrations also include a subsample of respondents for a reinterview. Reinterviews were conducted for ECE-NHES:1991; both SR-NHES:1993 and SS&D-NHES:1993; AE-NHES:1995; both PFI-NHES:1996 and YCI-NHES:1996; and ASPA-NHES:2001, AEWR-NHES:2003, and AE-NHES:2005.

In a reinterview, the respondent is asked to respond to the same items on different occasions. In order to limit

the response burden of the reinterview program, only selected items are included in the reinterview. The item selection criteria focus on the inclusion of key survey statistics (e.g., frequency of reading to children), items that are expected to have a potential for measurement error based on cognitive laboratory or field-test findings, and items required to control the question skip patterns for the reinterview. The results of the reinterviews are used to modify subsequent NHES surveys and to give some guidance to users about the reliability of responses for specific items in the data files. (See *Use of Cognitive Laboratories and Recorded Interviews in the National Household Education Survey* [Nolin 1997].) However, the reinterview procedure does not account for all measurement errors in the interviewing process, such as systematic errors that would be made in both the original interview and the reinterview.

The major emphasis of the 1991, 1993, and 1995 reinterview studies was to measure response variability. Overall, the results were positive. For example, within the AE-NHES:1995 reinterview study, only three items in one subject area had high response variability. The reinterview responses were consistent for most items; only minor modifications were suggested. (See *Measurement Error Studies at the National Center for Education Statistics* [Salvucci et al. 1997].)

Bias study. As part of the 2007 NHES administration a comprehensive bias study was conducted to look at the impact of non response and coverage issues on the NHES. The bias study utilized a separately drawn area probability sample and compared results to the RDD study. The study did not identify systematic patterns of bias in the key NHES statistics. However, some potential for bias was found in five estimates and concern over the ability of a landline frame to maintain adequate coverage in the future was raised. (See *An Evaluation of Bias in the 2007 National Household Education Surveys Program Results from a Special Data Collection Effort [Van de Kerckhove et. al. 2009]).*

Data Comparability

The NHES data can be compared with estimates from several other large-scale data collections, as described below.

Comparisons of methodology. For analysts wanting to compare the NHES surveys with another household survey, the Survey of Income and Program Participation (SIPP)—a longitudinal household survey conducted by the U.S. Bureau of the Census—provides an appropriate comparison. The first wave of data collection in SIPP is always done by personal visit to the household. Subsequent data collection is conducted primarily by telephone but may also be done in person. The response rates for SIPP are much higher than those that could be expected using an RDD screening sample, as in the NHES program. With personal interviews, there are more opportunities to obtain participation (including activities such as speaking with neighbors), and it is easier to demonstrate the importance of the sampled person's cooperation. It should be noted that, while the difference in response rates is largely the result of the different modes of sampling and data collection, the Census Bureau's response rates are generally higher than those achieved by other collection organizations.

Comparisons of topical data. Specific data from NHES surveys can be compared with data from several other surveys, as described below.

Early childhood education. Over the years, several NHES surveys have collected similar information on early childhood education: SR-NHES:2007, ECPP-NHES:2005, ECPP-NHES:2001, ECPP-NHES: 1995, ECE-NHES:1991, and SR-NHES:1993. These data can be compared with data from three other surveys. The CPS October Education Supplement collects information on nursery school enrollment. (See chapter 27.) CPS estimates of participation in early childhood programs and estimates of retention in early grades can be compared with NHES estimates. In addition, the 1990 CPS October Education Supplement replicated several NHES items on home activities that parents engage in with their children. NHES data can also be compared with data from the National Health Interview Survey Child Health Supplement of 1988 (conducted by the National Center for Health Statistics), which collected information on participation in child care and early childhood education programs and on the health status of children. Finally, SIPP (described above) periodically includes a supplement that collects information on the child care and early childhood program participation of children of mothers who are employed or enrolled in school or job training which is comparable with NHES data.

Before- and after-school programs and activities. PFI-NHES:2007 collected information on topics such as participation in literacy-related activities with family members, school size, contacts from the school, parent involvement with the school, disabling conditions, and parent and household characteristics. ASPA-NHES:2005 and ASPA-NHES:2001 covered some topics addressed in previous years by other NHES surveys. Parent-NHES:1999 and PFI/CI-NHES:1996 both collected information on school contacts with

households about children. Parent-NHES:1999 also collected information on type of care and basic statistics on after-school program participation. Basic enrollment totals and demographic characteristics, as well as public and private school enrollment data, from these NHES surveys, can be compared with CPS estimates.

Table 19. Weighted response rates for selected NHES surveys: 1991–2007

Questionnaire	Screener/1st stage	Interview/2nd stage	Overall
ECE-NHES:1991	81.0	94.5	76.5
AE-NHES:1991	81.0	84.7	68.6
SR-NHES:1993	82.1	89.6	73.6
SS&D-NHES:1993 – Parents, 3rd–5th	82.1	89.4	73.4
SS&D-NHES:1993 – Parents, 6th–12th	82.1	89.6	73.6
SS&D-NHES:1993 – Students, 6th–12th	82.1	83.0	68.1
ECPP-NHES:1995	73.3	90.4	66.3
AE-NHES:1995	73.3	80.0	58.6
PFI/CI-NHES:1996	69.9	89.4	62.5
YCI-NHES:1996	69.9	76.4	53.4
ACI-NHES:1996	69.9	84.1	58.9
Parent-NHES:1999	74.1	90.0	66.7
Youth-NHES:1999	74.1	78.1	57.9
AE-NHES:1999	74.1	84.1	62.3
AELL-NHES:2001	69.2	77.2	53.4
ECPP-NHES:2001	69.2	86.6	59.9
ASPA-NHES:2001	69.2	86.4	59.7
AEWR-NHES:2003	64.6	76.2	49.2
PFI-NHES:2003	64.6	83.3	53.8
AE-NHES:2005	66.9	71.2	47.6
ASPA-NHES:2005	66.9	84.1	56.3
ECPP-NHES:2005	66.9	84.4	56.4
AEWR-NHES:2007	52.8	62.4	33.0
PFI-NHES:2007	52.8	74.1	39.1
SR-NHES:2007	52.8	77.0	40.7

SOURCE: Brick, J.M., and Broene, P. (1997). *Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)* (NCES Working Paper 97-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC. Brick, J.M., Tubbs, E., Collins, M., and Nolin, M. (1997). *Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)* (NCES Working Paper 97-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC. Collins, M., Montaquila J., Nolin, M., Kim, K., Kleiner, B., and Waits, T. (2003). *National Household Education Surveys of 2001 Data File User's Manual, Volume I* (NCES 2003-079). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Montaquila, J., and Brick, J. M. (1997). *Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey* (NCES Working Paper 97-40). National Center for Education Statistics, U.S. Department of Education. Washington, DC. Nolin, M., Montaquila, J., Nicchitta, P., Kim, K., Kleiner, B., Lennon, J., Chapman, C., Creighton, S., and Bielick, S. (2000). *NHES:1999 Methodology Report* (NCES 2000-078). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Adult education. Both NHES surveys (AEWR-NHES:2003, AELL-NHES:2001, AE-NHES:1999, AE-NHES:1995, and AE-NHES:1991) and CPS provide estimates of adult education participation. (See chapter adult education that replicated items used to estimate the adult education participation rate in AE-NHES:1991.

School safety and discipline. Estimates from SS&D-NHES:1993 can be compared with estimates from three other surveys. The Monitoring the Future Survey (conducted annually by the National Institute on Drug Abuse) gathers information on the prevalence and incidence of the illicit drug use of 12th-graders. In addition, it contains questions designed to describe and explain changes in many important values, behaviors, and lifestyle orientations of American youth. The School Crime Supplement of the 1989 and 1995 National Crime Victimization Survey (conducted by the U.S. Department of Justice, Bureau of Justice Statistics) provides detailed information on personal crimes of violence and theft that were committed inside a school building or on school property. Finally, the NCES National Education Longitudinal Study of 1988 (NELS:88) provides data on educational issues such as the school environment, school discipline, victimization at school, and drug and alcohol education. (See chapter 8.)

Parent involvement in education. Estimates from PFI/CI-NHES:1996 can be compared with data from NELS:88. Data analysts may wish to examine NELS:88 data in conjunction with the PFI estimates on school contacts with parents (by parent report) and the frequency of parents helping their child with his or her homework.

Civic involvement and other characteristics. Estimates from the NHES Adult and Youth Civic Involvement Surveys can be compared with estimates from seven other surveys. The 1995 CPS October Education Supplement included sets of items measuring the percentage distribution of the adult population, age and sex of the adult population, household income distributions, and race/ethnicity by highest level of education. (See chapter 27.) The 1992 National Adult Literacy Survey collected data on adults' activities in daily life that require English literacy skills. (See chapter 19). Areas common to the 1994 General Social Survey, sponsored by the National Science Foundation, and ACI-NHES:1996 include organizational membership, various political or civic activities, and attitudes about freedom of speech. The National Election Study collects data on voting, public opinion, and political participation and knowledge during election years. Several items addressing political

27.) CPS collected information on adult education participation every 3 years from 1969 through 1984 The 1992 CPS also included a brief set of questions on

knowledge in ACI-NHES:1996 were drawn from the National Election Study and can be used for direct comparisons. The Citizens' Political and Social Participation Survey measures the extent and variety of voluntary social and political activity among Americans and the causes of that engagement. The Washington Post/Kaiser Family Foundation/Harvard University Survey Project provides information on public knowledge, perceptions, and attitudes about the role of American government. Finally, the National Survey of High School Seniors, a part of the CPS, elicits detailed information on political and relevant nonpolitical matters so that parent-child similarities and differences can be assessed. ACI-NHES:1999 expanded on the 1996 Youth Civic Involvement Survey by including more questions about youth service activities.

6. CONTACT INFORMATION

For content information on NHES, contact:

Andrew Zukerberg
Phone: (202) 219-7056
E-mail: andrew.zukerberg@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Collins, M., Montaquila, J., Nolin, M.J., Kim, K., Kleiner, B., Waits, T., Chapman, C., and Chandler, K. (2003). *National Household Education Surveys of 2001: Data File User's Manual, Volume I* (NCES 2003-079). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hagedorn, M., Montaquila, J., Carver, P., O'Donnell, K.,and Chapman, C. (2006). *National Household Education Surveys of 2005: Data File User's*

Manual (NCES 2006-078). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hagedorn, M., Montaquila, J., Vaden-Kiernan, N., Kim, K., and Chapman, C. (2004). *National Household Education Surveys of 2003: Data File User's Manual* (NCES 2004-101). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hagedorn, M., Roth, S.B., Carver, P., Van de Kerckhove, W., Smith, S., and Mulligan, G. (2009). *National Household Education Surveys Program of 2007: Methodology Report* (NCES 2009-047). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kim, K., and Brown, J. (1997). *The 1995 National Household Education Survey: Adult Education Course Code Merge Files User's Guide* (NCES Working Paper 97-20). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kim, K., and Collins, M.A. (1997). *The 1995 National Household Education Survey: Adult Education Course Coding Manual* (NCES Working Paper 97-19). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Nolin, M.J., Collins, M., and Brick, J.M. (1997). *An Overview of the National Household Education Survey: 1991, 1993, 1995, and 1996* (NCES 97-448). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Nolin, M.J., Montaquila, J., Nicchitta, P., Kim, K., Kleiner, B., Lennon, J., Chapman, C., Creighton, S., and Bielick, S. (2000). *National Household Education Survey of 1999: Methodology Report* (NCES 2000-078). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Survey Design

Brick, J.M. (1992). *Multiplicity Sampling for Dropouts in the NHES Field Test* (NCES 92-102). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., and Broene, P. (1997). *Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)* (NCES Working Paper 97-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Celebuski, C.A., and Collins, M.A. (1992). *Overview of the NHES Field Test* (NCES 92-099). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., and Collins, M. (1997). *An Overview of Response Rates in the National Household Education Survey: 1991, 1993, 1995, and 1996* (NCES 97-948). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., and Collins, M. (1997). *An Experiment in Random-Digit-Dial Screening* (NCES 98-255). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Collins, M., and Davies, E. (1997). *Feasibility of Conducting Follow-up Surveys in the National Household Education Survey* (NCES 97-335). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Collins, M.A., Nolin, M.J., Davies, E., and Feibus, M.L. (1997). *Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)* (NCES Working Paper 97-04). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Tubbs, E., Collins, M.A., and Nolin, M.J. (1997). *Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)* (NCES Working Paper 97-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., and Waksberg, J. (1991). Avoiding Sequential Sampling With Random Digit Dialing. *Survey Methodology, 17*(1): 27–42.

Casady, R.J., and Lepkowski, J.M. (1993). Stratified Telephone Survey Designs. *Survey Methodology, 19*(1): 103–113.

Collins, M., Brick, J.M., Loomis, L.S., Nicchitta, P.G., and Fleischman, S. (1997). *Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey (NHES:95)* (NCES Working Paper 97-08). National Center for

Education Statistics, U.S. Department of Education. Washington, DC.

Mohadjer, L. (1992). *Effectiveness of Oversampling Blacks and Hispanics in the NHES Field Test* (NCES 92-104). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Montaquila, J., and Brick, J.M. (1997). *Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey* (NCES Working Paper 97-40). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Vaden-Kiernan, N., Nicchitta, P.G., Montaquila, J., and Collins, M.A. (1997). *Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey* (NCES Working Paper 97-35). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Data Quality and Comparability

Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2009. National Center for Health Statistics. May 2010. Available from: http://www.cdc.gov/nchs/nhis.htm.

Blumberg, S.J., Luke, J.V., and Cynamon, M.L. (2006*)*. Telephone Coverage and Health Survey Estimates: Evaluating the Need for Concern about Wireless Substitution. *American Journal of Public Health*, *96*(5), 926-931.

Boyle, J.,Bucuvalas, M. Piekarski, L. and Weiss, A. (2009). Zero Banks Coverage Error and Bias in RDD Samples Based on Hundred Banks With List Assisted Numbers. *Public Opinion Quarterly*, *73*(4): 729-750.

Brick, J.M. (1996). *Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)* (NCES Working Paper 96-29). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., and Burke, J. (1992). *Telephone Undercoverage Bias of 14- to 21-Year-Olds and 3- to 5-Year-Olds* (NCES 92–201). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Keeter, S., Waksberg, J., and Bell, B. (1997). *National Household Education Survey of 1993: Adjusting for Coverage Bias Using Telephone Service Interruption Data* (NCES 97-336). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, M., Kim, K., Nolin, M.J., and Collins, M.A. (1996). *Estimation of Response Bias in the NHES:95 Adult Education Survey* (NCES Working Paper 96-13). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Rizzo, L., and Wernimont, J. (1997). *Reinterview Results for the School Safety and Discipline and School Readiness Components* (NCES Working Paper 97-339). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Tubbs, E., Collins, M.A., Nolin, M.J., Cantor, D., Levin, K., and Carnes, Y. (1997). *Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)* (NCES Working Paper 97-02). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Brick, J.M., Waksberg, J., Kulp, D., and Starer, A. (1995). Bias in List-Assisted Telephone Samples. *Public Opinion Quarterly*, *59*(2): 218–235.

Brick, J.M., Wernimont, J., and Montes, M. (1996). *The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component* (NCES Working Paper 96-14). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Collins, M.A., Branden, L., Nolin, M.J., Davies, E., and Loomis, L. (1997). *Comparison of Estimates From the 1993 National Household Education Survey (NHES:93)* (NCES Working Paper 97-34). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Collins, M.A., Brick, J.M., and Kim, K. (1997). *Measuring Participation in Adult Education* (NCES 97-341). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Collins, M.A., and Chandler, K. (1997). *A Guide to Using Data From the National Household Education Survey (NHES)* (NCES 97-561).

National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Fahimi, M., Kulp, D. and Brick, J.M. (2009). A Reassesment of List-Assisted RDD Methodology. *Public Opinion Quarterly, 73*(4): 751-760.

Hagedorn, M., Roth, S.B., Carver, P., Van de Kerckhove, W., and Smith, S. (2009). *National Household Education Surveys Program of 2007: Methodology Report.* (NCES 2009-047). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hagedorn, M., Roth, S.B., Carver, P., Van de Kerckhove, W., and Smith, S. (2009). *National Household Education Surveys Program of 2007: Methodology Report.* (NCES 2009-047). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kim, K., Loomis, L.S., and Collins, M.A. (1996). *Comparison of Estimates From the 1995 National Household Education* Survey *(NHES:95)* (NCES Working Paper 96-30). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Mohadjer, L., and Brick, J.M. (1992). *Proxy Reporting of Dropout Status in the NHES Field Test* (NCES 92-103). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Montaquilla, J.M., Brick, J.M., and Brock, S.P. (1997). *Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey* (NCES Working Paper 97-38). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Montaquilla, J.M., Brick, J.M., and Brock, S.P. (1997). *Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey* (NCES Working Paper 97-39). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Nolin, M.J. (1997). *Use of Cognitive Laboratories and Recorded Interviews in the National Household Education Survey (NHES)* (NCES 96-332). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Tucker, C., Lepkowski, J.M., and Piekarski, L. (2002). The Current Efficiency of List-Assisted Telephone Sampling Designs. *Public Opinion Quarterly, 66*(3): 321-338.

Van de Kerckhove, W., Montaquila, J.M.,Carver, P.R., and J.M. Brick. (2009). *An Evaluation of Bias in the 2007 National Household Education Surveys Program Results from a Special Data Collection Effort* (NCES 2009-0029). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 27: Current Population Survey (CPS) – October Supplement

1. OVERVIEW

The Current Population Survey (CPS) is a monthly survey of 50,000–60,000 households conducted by the Bureau of the Census, part of the U.S. Department of Commerce, for the Bureau of Labor Statistics (BLS), U.S. Department of Labor. The basic monthly CPS collects data about the employment, unemployment, and other characteristics of the civilian noninstitutionalized population in the United States; it excludes military personnel and their families living on post, inmates of institutions, and residents of homes for the aged. Since the late 1960s, the National Center for Education Statistics (NCES) has sponsored the October Supplement to the CPS to capture additional information on school enrollment status and related topics for household members 3 years old and over, thus providing current estimates of school enrollment as well as of the social and economic characteristics of students.

SUPPLEMENTS TO THE CPS

September Supplement:

- Conducted only in 2001
- Collected data on the availability and use of computers and the Internet at school, home, and work.

October Supplement:

- Conducted annually
- Collects data on household members 3 years old and over on school enrollment status.

Purpose

The October Supplement is designed to collect information on the school enrollment of household members in any type of public, parochial, or other private school in the regular school system. Such schools include nursery schools, kindergartens, elementary schools, high schools, colleges, universities, and professional schools. Additional supplementary questions are designed to collect information on various topics of interest.

Components

The October Supplement is an annual addition to the basic monthly CPS. The information collected is described below. A member of each household who is at least 15 years old provides information for all members of the household.

October Supplement. The October Supplement collects information on the school enrollment status and educational attainment of household members 3 years old and over, including highest grade completed, level and grade of current enrollment, attendance status, number and type of courses taken, degree or certificate objective, and type of organization offering instruction for each member of the household. A dozen core questions in the interview instrument for the October Supplement have remained unchanged since 1967. Since 1987, additional questions have been included on business, vocational, technical, secretarial, trade, and correspondence courses; on the grade the student was attending in the previous year; on the calendar year that the student received his or her most recent degree; on whether or not the student completed high school by means of an equivalency test (such as a General Educational Development [GED] credential); and on whether or not children ages 3 to 5 are enrolled in any kind of nursery school, kindergarten, or elementary school. From time to time, additional items address such topics as private school tuition, adult education,

vocational education, computer and internet use, language proficiency library,use, disability status, and student mobility.

Basic CPS. The basic CPS collects monthly data on household membership, household characteristics, demographic characteristics, and labor force participation of the civilian noninstitutionalized population 15 years of age and over. However, published data focus on those ages 16 and over. The basic CPS is collected each month from a probability sample of approximately 50,000–60,000 occupied households.

Periodicity

The basic CPS is conducted monthly. The October Supplement to the CPS is an annual supplement.

2. USES OF DATA

The October Supplement provides important education data to policymakers and researchers on school enrollment and educational attainment. Data from the October Supplement, together with data from the basic CPS and the March Supplement (Annual Social and Economic Supplement), provide the basis for descriptive and analytic reports that portray the social and economic characteristics of students in relation to the specifics of their school enrollment. From these sources, it is possible to retention rates and completion rates for various levels of education, and high school dropout. In some years, the October Supplement also provides policy-relevant data on private school tuition, adult education, vocational education, early childhood education, and student mobility.

3. KEY CONCEPTS

Some of the key concepts in the CPS October Supplement are defined below. For additional terms relevant to the October Supplement, as well as to the basic CPS, refer to *School Enrollment—Social and Economic Characteristics of Students: October 1996 (Update). Detailed Tables and Documentation for P20-500* (U.S. Department of Commerce 1998).

Household. All persons who occupy a housing unit. A house, an apartment or other group of rooms, or a single room is regarded as a housing unit when it is occupied or intended for occupancy as separate living quarters; that is, when the occupants do not live and eat with any other persons in the structure and there is direct access from the outside or through a common hall. A household includes the related family members and all the unrelated persons, if any, such as lodgers, foster children, wards, or employees who share the housing unit. A person living alone in a housing unit, or a group of unrelated persons sharing a housing unit as partners, is also counted as a household.

School Enrollment. School enrollment includes anyone who has been enrolled at any time during the current term or school year in any type of public, parochial, or other private school in the regular school system. Such schools include nursery schools, kindergartens, elementary schools, high schools, colleges, universities, and professional schools. Attendance may be either full time or part time, during the day or night. Regular schooling is that which may advance a person toward an elementary or high school diploma, or a college, university, or professional school degree. Enrollment is excluded if in schools that are not in the regular school system or that do not advance students to regular school degrees (e.g., enrollment in trade schools, business colleges, and schools for the mentally handicapped).

Level of School. Nursery school, kindergarten, elementary school (1^{st} through 8^{th} grades), high school (9^{th} through 12^{th} grades), and college or professional school. The last level includes graduate students in colleges or universities. Persons enrolled in elementary school, middle school, intermediate school, or junior high school through the 8^{th} grade are classified as in elementary school. All persons enrolled in the 9^{th} through 12^{th} grades are classified as in high school.

Nursery School. A group or class that is organized to provide educational experiences for children during the year or years preceding kindergarten. This includes Head Start programs or similar programs sponsored by local agencies to provide preschool education to young children.

Public or Private School. A public school is defined as any educational institution operated by publicly elected or appointed school officials and supported by public funds. Private schools include educational institutions established and operated by religious bodies, as well as those that are under other private control. In cases where enrollment is in a school or college that is both publicly and privately controlled or supported, enrollment is counted according to whether it is primarily public or private.

Modal Grade. For descriptive and analytic purposes, enrolled persons are classified according to their relative progress in school; that is, whether the grade or year in which they were enrolled was below, at, or above the modal (or typical) grade for persons of their age at the time of the survey. The modal grade is the year of school in which the largest proportion of students of a given age are enrolled.

Vocational School Enrollment. Vocational school enrollment includes enrollment in business, vocational, technical, secretarial, trade, and correspondence courses not counted as regular school enrollment and not for recreation or adult education classes.

Educational Attainment. Highest level of school a person has completed or highest degree a person has received.

4. SURVEY DESIGN

Target Population

All household members age 3 and older in the civilian noninstitutionalized population of the 50 states and the District of Columbia. Excludes military personnel and their families living on post, inmates of institutions, and residents of homes for the aged.

Sample Design

The CPS sample is a multistage stratified sample of approximately 72,000 assigned housing units from 824 sample areas designed to measure the demographic and labor force characteristics of the civilian noninstitutionalized population 15 years of age and older. Published data, however, focus on those ages 16 and over. Currently, the CPS samples housing units from lists of addresses obtained from the 2000 Decennial Census of Population and Housing. The sample is updated continuously for new housing built after the 2000 Census.

To improve the reliability of estimates of month-to-month and year-to-year change, eight panels of housing units are used to rotate the sample each month. A sample unit is interviewed for 4 consecutive months and then, after an 8-month rest period, for the same 4 months a year later. Every month, a new panel of housing units, or one-eighth of the total sample, is introduced. Thus, in a particular month, one panel is being interviewed for the first time, one panel for the second, and so on.

The first-stage sample selection is carried out in three major steps: definition of the primary sampling units (PSUs), stratification of the PSUs within each state, and selection of the sample PSUs in each state. There are currently (after the 2000 Decennial Census) 2,025 defined PSUs in the United States from which to draw the CPS sample. The CPS sample design calls for combining PSUs into strata within each state and selecting one PSU from each stratum. The CPS currently uses the Stratification Search Program (SSP), created by the Demographic Statistical Methods Division of the Census Bureau, to perform the PSU stratification. CPS strata in all states except Alaska are formed using the SSP. (A separate program performs the stratification for Alaska.) A total of 824 PSUs are selected for the sample. Using a procedure designed to maximize overlap, one PSU is selected per stratum with probability proportional to its 2000 population. This procedure uses mathematical programming techniques to maximize the probability of selecting PSUs that are already in sample while maintaining the correct overall probabilities of selection.

The second stage of the CPS sample design is the selection of sample housing units within PSUs. These ultimate sampling unit (USU) clusters consist of a geographically compact cluster of approximately four addresses, corresponding to four housing units at the time of the census. Each month, about 72,000 housing units are assigned for data collection, of which about 60,000 are occupied and thus eligible for interview. The remainder are units found to be destroyed, vacant, converted to nonresidential use, containing persons whose usual place of residence is elsewhere, or ineligible for other reasons. Of the 60,000 housing units, about 5 percent are not interviewed in a given month due to temporary absence (vacation, etc.), other failures to make contact after repeated attempts, the inability of persons contacted to respond, unavailability for other reasons, and refusals to cooperate (which make up about half of the noninterviews). Information is obtained each month on for approximately 110,000 persons 15 years of age or older and on approximately 30,000 persons under the age of 15.

Since 2005, the CPS sample has been selected based on 2000 census information. From 1995 to 2004, the sample was based on 1990 census information; samples prior to 1995 similarly used earlier censuses. The number of PSUs, housing units, and persons interviewed are also different in samples prior to 2005. Specifics on each given CPS sample can be found in the technical documentation report for the year’s CPS.

Data Collection and Processing

The U.S. Bureau of the Census is the collection agent for the CPS and its supplements. Additional details on data collection and processing are provided in *The Current Population Survey: Design and Methodology* (Technical Paper 66) (U.S. Department of Commerce 2006).

Reference Dates. The reference period for the October Supplement is the current school year, which is assumed to be in progress in the interview month of October. The CPS labor force questions ask about labor market activities for 1 week each month. This week is referred to as the ''reference week.'' The reference week is defined as the 7-day period, Sunday through Saturday, which includes the 12th of the month.

Data Collection. Each month, Bureau of the Census field representatives attempt to collect data from the sample units during the week containing the 19th of the month. For the first month-in-sample interview, the interviewer visits the sample address to determine if the sample unit exists, if it is occupied, and if some responsible adult will provide the necessary information. If someone at the sample unit agrees to the interview, the interviewer uses a laptop computer to administer the interview. In most cases, the interviewer conducts subsequent interviews by telephone (use of telephone interviewing must be approved by the respondent) and does not actually visit the sample unit again until the fifth month-in-sample interview, the first interview after the 8-month resting period. Fifth-month households are more likely than any other household to be a replacement household; that is, a household in which all the previous month's residents have moved out and been replaced by an entirely different group of residents. However, any person can change his or her household status during the time in sample: a person who leaves the household is deleted from the roster; a person who moves into the household is added to the roster.

Most month-in-sample 2 through 4 and 6 through 8 interviews are conducted by telephone. (For instance, 78.8 percent of the interviews for the October 2004 Supplement were conducted by telephone, which is highly consistent with the usual monthly results for telephone interviews.) Interviewers continue to visit households without telephones, with poor English language skills, or that decline a telephone interview.

The interview begins with questions about the housing unit and the people who consider this address their usual residence. Basic demographic information is collected for each household member. Labor force information is collected for each civilian 15 years of age or older, although the data for 15-year-olds are not used in official BLS estimates. After the labor force information has been collected for all eligible household members, supplemental questions particular to that month's interview may be asked of specific family members or the entire household.

Editing. Completed interviews are electronically transmitted to a central processor where the responses are edited for consistency and various codes are added. The edits effectively blank out all entries in inappropriate questions and ensure that all appropriate questions have valid entries.

Estimation Methods

Weighting is used in the CPS to adjust for sampling and unit nonresponse, and imputation is used to adjust for item nonresponse.

Weighting. For the basic CPS, the estimation procedure involves weighting the data from each sample person by the inverse of the probability of the person's housing unit being in the sample. With some exceptions, sample persons within the same state have the same probability of selection. The CPS uses raking ratio estimation to derive the weights used to tabulate total U.S. and state estimates. The goal is to control the survey estimates of the population in specific subgroups to match independently obtained estimates of the civilian noninstitutionalized population in the 50 states and the District of Columbia. These population estimates are prepared monthly to agree with the most current set of population estimates that are released as part of the Census Bureau's population estimates and projections program. In addition, household and family weights provide a basis for household-level estimates and estimates for married couples living in the same household.

For all CPS data files, a final weight is prepared and used to compute the monthly labor force status estimates. The final weight, which is the product of several adjustments, including a nonresponse adjustment, is used to produce estimates for the various characteristics covered in the full monthly CPS. This weight is constructed from the basic weight for each person, which represents the probability of selection for the survey. For supplements, such as the October Supplement, separate data processing is required, not only to edit responses for consistency and impute for missing values, but also to incorporate special weighting procedures to account for the fact that the

supplement is targeting a special universe, such as school-age children, in contrast to the working-age labor force emphasis of the basic CPS.

Starting with the data collected in the October 1994 CPS, independent estimates have been based on civilian noninstitutionalized population controls for age, race, and sex established by the decennial census and adjusted to compensate for an undercount. These independent estimates are based on statistics from decennial censuses; statistics on births, deaths, immigration, and emigration; and statistics on the size of the Armed Forces.

Imputation. When a response is not obtained for a particular data item, or an inconsistency in reported items is detected, an imputed response is entered in the field. Before the edits are applied, the daily data files are merged and the combined file is sorted by state and PSU within state. This sort ensures that allocated values are from geographically related records; that is, missing values for records in Maryland will not receive values from records in California. This is an important distinction since many labor force and industry and occupation characteristics are geographically clustered. The edits are run in a deliberate and logical sequence. Demographic variables are edited first because several of these variables are used to allocate missing values in the other modules. The labor force module is edited next, since labor force status and related items are used to impute missing values for industry and occupation codes and so forth.

CPS edits use three imputation methods: relational imputation, longitudinal edits, and hot-deck imputation. Relational imputation infers the missing value from other characteristics in the person's record or within the household. Longitudinal edits are used primarily in the labor force edits. If a question is blank and the record is in the overlap sample, the edit looks at the previous month's data to determine whether the person had responded then for that item. If so, the previous month's entry is assigned; otherwise, the item is assigned a value using the appropriate hot deck. The hot-deck method assigns a value from a record with similar characteristics. Hot decks are always defined by age, race, and sex. Other characteristics used in hot decks vary depending on the nature of the question being referenced. The imputation procedure is performed one item at a time. In a typical month, the imputation rate for demographic items is less than 1 percent. The rates for labor force items are slightly over 1 percent. Over all earnings items, the imputation rate is near 10 percent, with some items having much higher and others much lower nonresponse rates. In October 2005, the imputation rate for the basic school enrollment items ranged from 4 to 7 percent per item.

Future Plans

The October Supplement will always include the traditional school enrollment questions; questions on other topics will be added as occasion warrants. For example, over the last several decades NCES has funded additional items on education-related topics such as language proficiency, disabilities, computer use and access, student mobility, and private school tuition. Plans for additional questions in future years have yet to be determined.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Although the estimation methods used in the CPS do not produce unbiased estimates, biases for most estimates are believed to be small enough so that the confidence interval statements are approximately true. Standard error estimates are computed using replicate variance techniques and reflect contributions not only from sampling error but also from some types of nonsampling error, particularly response variability and intra-interviewer correlation. Because replicate variance techniques are somewhat cumbersome, simplified formulas called generalized variance functions (GVFs) have been developed for various types of labor force characteristics. The GVF can be used to approximate an estimate's standard error, but this only indicates the general magnitude of its standard error rather than a precise value. Standard error estimates computed using generalized variance functions are provided in *Employment and Earnings* and other BLS publications.

Nonsampling Error

Although the full extent of nonsampling error in the CPS is unknown, special studies have been conducted to quantify some of the possible sources. The effect of nonsampling error should be small on estimates of relative change, such as month-to-month change. Estimates of monthly levels would be more severely affected by nonsampling error.

Coverage Error. The concept of coverage in the survey sampling process is the extent to which the total population that could be selected for the sample "covers" the survey's target population. Undercoverage

in the CPS results from missed housing units and missed persons within sample households. Overall CPS undercoverage for households was estimated to be about 10 percent for October 2005 and about 11 percent for October 2006. It is known that the CPS undercoverage varies with age, sex, race, and Hispanic origin. Generally, undercoverage is larger for men than for women and larger for Blacks, Hispanics, and other races than for Whites. Ratio adjustment to independent age/sex/race/origin population controls, as described previously, partially corrects for the biases due to survey undercoverage. However, biases exist in the estimates to the extent that missed persons in missed households or missed persons in interviewed households have different characteristics than interviewed persons in the same age/sex/race/origin group.

The independent population estimates used in the estimation procedure may be a source of error although, on balance, their use substantially improves the statistical reliability of many of the figures. Errors may arise in the independent population estimates because of underenumeration of certain population groups or errors in age reporting in the decennial census (which serves as the base for the estimates) or similar problems in the components of population change (mortality, immigration, etc.).

Nonresponse Error.

Unit Nonresponse. Unit nonresponse may have a number of components. A respondent may refuse to participate in the survey, may not be capable of completing the interview, or may not be available to the interviewer during the specified survey period. If the entire household does not participate, this situation is referred to as a "Type A noninterview." There is also another type of (partial) unit nonresponse, namely, that one or more individual persons within the household refuse to be interviewed. This is not a major problem in the CPS since any responsible adult may be able to report information for other persons as a proxy reporter. There are other variations on unit nonresponse; detailed consideration of these may be found in *The Current Population Survey: Design and Methodology* (Technical Paper 66) (U.S. Department of Commerce 2006).

For the October 2005 basic CPS, the nonresponse rate was 7.4 percent, and the nonresponse rate for the October supplement was an additional 3.4 percent. These two nonresponse rates led to a combined nonresponse rate of 10.5 percent. For the October 2006 basic CPS, the household-level nonresponse rate was 8.1 percent, and the person-level nonresponse rate for the October supplement was an additional 3.9 percent. Since the basic CPS nonresponse rate was a household-level rate and the School Enrollment supplement nonresponse rate was a person-level rate, these rates couldn't be combined to derive an overall nonresponse rate. Since it is unlikely the nonresponding households to the basic CPS had the same number of persons as the households successfully interviewed, combining these rates would have resulted in an overestimate of the "true" person-level overall nonresponse rate for the October supplement (for more information, see *The Current Population Survey October 2006: School Enrollment Supplement Technical Documentation,* U.S. Department of Commerce 2006).

Item Nonresponse. Although an imputation procedure is implemented for item nonresponse in the CPS, there is no way of ensuring that the errors of item imputation will balance out and that any potential bias has been avoided.

Measurement Error. The main sources of nonsampling variability in the responses to the October Supplement are those inherent in the survey instrument. The question of current school enrollment may not be answered accurately for various reasons. Some respondents may not know current grade information for every student in the household, a problem especially prevalent for households with members in college or in nursery school. Confusion over college credits or hours taken by a student may make it difficult to determine the year in which the student is enrolled. Problems may occur with the definition of nursery school (a group or class organized to provide educational experiences for children), where respondents' interpretations of "educational experiences" vary.

Data Comparability

NCES collects preschool, elementary school, secondary school, and postsecondary education enrollment and completion data through a wide range of studies including the National Household Education Surveys Program (NHES, see chapter 26), the Common Core of Data (CCD, see chapter 2), the Private School Survey (PSS, see chapter 3), the Integrated Postsecondary Education Data System (IPEDS, see chapter 12), and the National Postsecondary Student Aid Study (NPSAS, see chapter 14). In addition, the Bureau of the Census collects the American Community Survey (ACS), which is another household survey that includes some school enrollment and educational attainment data.

Because of differences in data collection modes, respondent selection, interviewer training, collection and reference periods, and differing survey processes, data obtained from the CPS and other sources are not entirely comparable. This is an example of nonsampling variability that is not reflected in the standard errors. Therefore, caution should be used when comparing results from different sources.

6. CONTACT INFORMATION

For content information about the basic monthly CPS and October Supplement, contact:

NCES Contact:
Chris Chapman
Phone: (202) 502-7414
E-mail: chris.chapman@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

Census Bureau Contact:
Lisa Clement
Phone: (301) 763-5482
E-mail: lisa.a.clement@census.gov

Mailing Address:
Education and Social Stratification
Branch, Population Division
Bureau of the Census
U.S. Department of Commerce
Washington, DC 20233

7. METHODOLOGY AND EVALUATION REPORTS

General

Kominski, R. (1987). *School Enrollment—Social and Economic Characteristics of Students*. Current Population Reports, P20-413. U.S. Department of Commerce. Washington, DC: Bureau of the Census.

Steinberg, J., Jabine, T.B., and Pritzker, L. (1963). *The Current Population Survey: A Report on Methodology*, Technical Paper 7. U.S. Department of Commerce. Washington, DC: Bureau of the Census.

U.S. Department of Commerce, Bureau of the Census. (1998). *School Enrollment—Social and Economic Characteristics of Students: October 1996 (Update). Detailed Tables and Documentation for P20-500*. U.S. Department of Commerce. Washington, DC: Bureau of the Census.

U.S. Department of Commerce, Bureau of the Census. (2005). *The Current Population Survey October 2005: School Enrollment Supplement Technical Documentation*. U.S. Department of Commerce. Washington, DC: Bureau of the Census.

U.S. Department of Commerce, Bureau of the Census. (2006). *The Current Population Survey October 2006: School Enrollment Supplement Technical Documentation*. U.S. Department of Commerce. Washington, DC: Bureau of the Census.

U.S. Department of Commerce, Bureau of the Census, and U.S. Department of Labor, Bureau of Labor Statistics. (1976). *Concepts and Methods Used in Labor Force Statistics Derived From the Current Population Survey*, Current Population Reports, Special Studies Series P23, No. 62. Washington, DC: Author.

U.S. Department of Labor, Bureau of Labor Statistics. (1997). *BLS Handbook of Methods*. Washington, DC: Author.

Survey Design

Frankel, L.R., and Stock, J.S. (1942). On the Sample Survey of Unemployment. *Journal of the American Statistical Association, 37:* 77–80.

Hansen, M.H., and Hurwitz, W.N. (1943). On the Theory of Sampling From Finite Populations. *Annals of Mathematical Statistics*, *XIV*: 333–362.

Shiskin, J., Young, A.H., and Musgrave, J.C. (1965). *The X11 Variant of the Census Method I Seasonal Adjustment Program*, Technical Paper 15. U.S. Department of Commerce. Washington, DC: Bureau of the Census.

U.S. Department of Commerce, Bureau of the Census. (2002). *The Current Population Survey: Design and Methodology*, Technical Paper 63RV. Washington, DC: Author.

U.S. Department of Commerce, Bureau of the Census. (2006). *The Current Population Survey: Design and Methodology*, Technical Paper 66. Washington, DC: Author.

Data Quality and Comparability

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory Vol. I, Methods and Applications.* New York: John Wiley.

Hansen, M.H., Hurwitz, W.N., Nisselson, H., and Steinberg, J. (1955). The Redesign of the Census Current Population Survey. *Journal of the American Statistical Association, 50*: 701–719.

U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. (1978). *An Error Profile: Employment as Measured by the Current Population Survey*, Statistical Policy Working Paper 3. Washington, DC: Author.

U.S. Department of Commerce, Bureau of the Census. (1978). *The Current Population Survey: Design and Methodology,* Technical Paper 40. Washington, DC: Author.

Chapter 28: Crime and Safety Surveys

TWO CRIME AND SAFETY SURVEYS:

- School Crime Supplement
- School Survey on Crime and Safety

The National Center for Education Statistics (NCES) conducts two surveys on a regular basis to collect data on school crime and safety: the School Crime Supplement (SCS) to the National Crime Victimization Survey (NCVS), a survey of students ages 12 through 18; and the School Survey on Crime and Safety (SSOCS), a survey of public schools and principals.

1. SCHOOL CRIME SUPPLEMENT (SCS)

Overview

The SCS is conducted on a biennial basis as a supplement to the NCVS, which is administered by the Bureau of Justice Statistics (BJS), U.S. Department of Justice, and conducted by the U.S. Census Bureau. The NCVS is an ongoing household survey that gathers information on the criminal victimization of household members age 12 and older. NCES and BJS jointly created the SCS to study the relationship between victimization at school and the school environment.

The SCS is designed to assist policymakers—as well as academic researchers and practitioners at the federal, state, and local levels—in making informed decisions concerning crime in schools. The SCS gathers data from nationally representative samples of students who are between the ages of 12 and 18 and who are enrolled in grades 6–12 in U.S. public or private schools. Prior to 2007, eligible sample members were those who had attended school at any time during the 6 months preceding the interview. In 2007, the questionnaire was changed to include students who attend school at any time during the school year.

The SCS asks students a number of questions about their experiences with, and perceptions of, crime and violence occurring inside their school, on school grounds, on the school bus, and from 2001 onward, going to or from school. The SCS contains questions not included in the NCVS, such as those on preventive measures employed by schools; students' participation in after-school activities; students' perceptions of school rules and the enforcement of these rules; the presence of weapons, drugs, alcohol, and gangs in school; student bullying and cyber-bullying; hate-related incidents; and students' attitudes related to the fear of victimization at school. The SCS was conducted in 1989, 1995, 1999, 2001, 2003, 2005, 2007, and 2009. Future administrations are planned at 2-year intervals in odd-numbered years.

Sample Design

Each month, the U.S. Census Bureau selects respondents for the NCVS using a "rotating panel" design. Households are selected into the sample using a stratified, multistage cluster design. In the first stage, the primary sampling units (PSUs), consisting of counties or groups of counties, are selected, and smaller areas, called Enumeration Districts (EDs), are selected within each sampled PSU. Large PSUs are included in the sample automatically and are considered to be self-representing strata since all of them are selected. The remaining PSUs (called non-self-representing because only a subset of them are selected) are combined into strata by grouping PSUs with similar geographic and demographic characteristics, as determined by the decennial census. Within each ED, clusters of four households, called segments, are selected. Across all EDs, sampled households are then divided into discrete groups (rotations), and all age-eligible individuals in the households

become part of the panel. Such a design ensures a self-weighting probability sample of housing units and group-quarter dwellings within each of the selected areas. (“Self-weighting” means that prior to any weighting adjustments, each sample housing unit had the same overall probability of being selected.)

To account for units built within each of the sample areas after the decennial census, a sample of permits issued for the construction of residential housing is drawn. Jurisdictions that do not issue building permits are sampled using small land-area segments. These supplementary procedures, though yielding a relatively small portion of the sample, enable persons living in housing units built after the decennial census to be properly represented.

In order to conduct field interviews for the NCVS, the sample of households is divided into six groups, or rotations. Each group of households is interviewed seven times—once every 6 months over a period of 3 years. Each rotation group is further divided into six panels. A different panel of households, corresponding to one-sixth of each rotation group, is interviewed each month during the 6-month period. Because the NCVS is continuous, newly constructed housing units are selected as described above, and assigned to rotation groups and panels for subsequent incorporation into the sample. A new rotation group enters the sample every 6 months, replacing a group phased out after 3 years. This type of rotation scheme is used to reduce the respondent burden that might result if households were to remain in the sample permanently. It should be noted that the data from the NCVS/SCS interviews obtained in the incoming rotation are included in the SCS data files. The incoming rotation was included in the NCVS data file only in 2007.

Once in the panel, NCVS interviews are conducted with all household members age 12 or older. After completion of the NCVS interview, an SCS interview is given to eligible household members. In order to be eligible for the SCS, students must be 12 through 18 years old, have attended school in grades 6 through 12 at some point during the school year, and not have been homeschooled during the school year. Persons who have dropped out of school, have been expelled or suspended from school, or are temporarily absent from school for any other reason, such as illness or vacation, are eligible as long as they attended school at any time during the school year. For the 1989 and 1995 SCS, 19-year-old household members were considered eligible for the SCS interview. Prior to the 2007 SCS, household members who were enrolled in school sometime during the previous 6 months prior to the interview were eligible.

Data Collection and Processing

In all SCS survey years, the SCS was conducted for a 6-month period from January through June in all households selected for the NCVS. Eligible respondents were asked the supplemental questions in the SCS only after completing their entire NCVS interview.

The 2007 SCS was fully automated; all interviews were conducted through computer-assisted personal interviewing (CAPI), where field representatives used questionnaires loaded into laptop computers to conduct interviews, which could be completed either in person (for the first and subsequent interviews, as circumstances called for) or by telephone. Two modes of data collection were used through the 2005 collection: (1) paper-and-pencil interviewing, which was conducted in person for the first NCVS/SCS interview; and (2) computer-assisted telephone interviewing (CATI), unless circumstances called for an in-person interview. There were 5,620 students who participated in the SCS in 2007; 6,300 in 2005; 7,150 in 2003; 8,370 in 2001; 8,400 in 1999; 9,730 in 1995; and 10,450 in 1989. The 2009 data have been collected but not yet released.

Interviewers are instructed to conduct interviews in privacy unless respondents specifically agree to permit others to be present. Most interviews are conducted over the telephone, and most questions require “yes” or “no” answers, thereby affording respondents a further measure of privacy. While efforts are made to assure that interviews about student experiences at school are conducted with the students themselves, interviews with proxy respondents are accepted under certain circumstances. These include interviews scheduled with a child between the ages of 12 and 13 where parents refuse to allow an interview with the child; interviews where the subject child is unavailable during the period of data collection; and interviews where the child is physically or emotionally unable to answer for him- or herself.

Weighting

The purpose of the SCS is to be able to make inferences about criminal victimization in the 12- to 18-year-old student population in the United States. Before such inferences can be drawn, it is important to adjust, or “weight,” the sample of students to ensure it is similar to the entire population in this age group. The SCS weights are a combination of household-level and person-level adjustment factors. In the NCVS, adjustments are made to account for both household- and person-level noninterviews. Additional factors are then applied to reduce the variance of the estimate by correcting for the differences between the sample

distributions of age, race, and sex and the known population distributions of these characteristics. The resulting weights are assigned to all interviewed households and persons in the file.

A special weighting adjustment is then made for the SCS respondents, and noninterview adjustment factors are computed to adjust for SCS interview nonresponse. This noninterview factor is applied to the NCVS person-level weight for each SCS respondent. Prior to 2007, two weights were available in the SCS data file. The first SCS weight was to be used if producing NCVS estimates using only the continuing rotations. The second SCS weight was derived using the final NCVS person weight that was calculated for all interviewed persons in continuing and incoming rotations. In 2007, all rotations were used for both the SCS and NCVS.

Table 20. Unweighted household, student, and overall unit response rates for the School Crime Survey: 1989–2007

Year	Household response rate	Student response rate	Overall response rate
1989	97	87	84
1995	95	78	74
1999	94	78	73
2001	93	77	72
2003	92	70	64
2005	91	62	56
2007	90	58	53

SOURCE: Chandler, K.A., Chapman, C.D., Rand, M.R., and Taylor, B.M. (1998). *Students' Reports of School Crime: 1989 and 1995* (NCES 98-241/NCJ-169607). National Center for Education Statistics, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC. Dinkes, R., Kemp, J., and Baum, K. (2009). *Indicators of School Crime and Safety: 2008* (NCES 2009-022/NCJ-226343). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Imputation

Item response rates are generally high. Most items are answered by over 95 percent of all eligible respondents. No explicit imputation procedure is used to correct for item nonresponse.

Sampling Error

Standard errors of percentage and population counts are calculated using the Taylor series approximation method using PSU and strata variables from the 1995, 1999, 2001, 2003, 2005, and 2007 data sets.

Another way in which the standard errors can be calculated, and were calculated in 1989, is by using the generalized variance function (GVF) constant parameters. The GVF represents the curve fitted to the individual standard errors that are calculated using the jackknife repeated replication technique.

Coverage Error

The decennial census is used for sampling housing units in the NCVS. To account for units built since the census was taken, supplemental procedures are implemented. (See "Sample Design" above.) Coverage error in the NCVS (and SCS), if any, would result from coverage error in the census and the supplemental procedures.

Unit Nonresponse

Because interviews with students can only be completed after households have responded to the NCVS, the unit completion rate for the SCS reflects both the household interview completion rate and the student interview completion rate (see table 20). Thus, the overall unweighted SCS response rate is calculated by multiplying the household completion rate by the student completion rate.

Due to the low student response rates in 2005 and 2007, unit nonresponse bias analyses were commissioned. In 2007, the analysis of unit nonresponse bias found evidence of bias by race, household income, and urbanicity variables. Hispanic respondents had lower response rates than respondents from other races/ethnicities. Respondents from households with an income of $25,000 or more had higher response rates than those from households with incomes of less than $7,500. Respondents who live in urban areas had lower response rates than those who live in rural areas. However, when responding students were compared to the eligible NCVS sample, there were no measurable differences between the responding students and the eligible students, suggesting the nonresponse bias has little impact on the overall estimates.

The analysis of unit nonresponse bias in 2005 also found evidence of bias for the race, household income, and urbanicity variables. White, non-Hispanic and other, non-Hispanic respondents had higher response rates than Black, non-Hispanic and Hispanic respondents. Respondents from households with incomes of $35,000–49,999 and $50,000 or more had higher response rates than those from households with incomes of less than $7,500, $7,500–14,999, $15,000–

24,999, and $25,000–34,999. Respondents who live in urban areas had lower response rates than those who live in rural or suburban areas.

Item Nonresponse

Item response rates for the SCS have been high. In all administrations, most items were answered by over 95 percent of all eligible respondents, with a few exceptions. One notable exception was the household income question, which was answered by about 80 percent of all households in 2007; about 82 percent of all households in 2005; and about 83, 84, 86, 90, and 90 percent of all households in 2003, 2001, 1999, 1995, and 1989, respectively. Due to their sensitive nature, income and income-related questions typically have relatively lower response rates than other items.

Measurement Error

Measurement error can result from respondents' different understandings of what constitutes a crime, memory lapses, and reluctance or refusal to report incidents of victimization. A change in the screener procedure between 1989 and 1995 was designed to result in the reporting of more incidents of victimization, more detail on the types of crime, and presumably more accurate data in 1995 than in 1989. (See "Data Comparability" below for further explanation.) Differences in the questions asked in the NCVS and SCS, as well as the sequencing of questions (SCS after NCVS), might have also lead to better recall in the SCS in 1995.

Data Comparability

The SCS questionnaire has been modified in several ways since its inception, as has the larger NCVS. Users making comparisons of data across years should be aware of the changes detailed below and their impact on data comparability. In 1989 and 1995, respondents to the SCS were asked two separate sets of questions regarding personal victimization. The first set of questions was part of the main NCVS, and the second set was part of the SCS. When examining data from either 1989 or 1995, the following have an impact on the comparability of data on victimization: (1) differences between years in the wording of victimization items in the NCVS as well as the SCS questionnaires; and (2) differences between SCS and NCVS items collecting similar data.

NCVS design changes. The NCVS was redesigned in 1992. Changes to the NCVS screening procedure put in place in 1992 make comparisons of 1989 data with those from later years difficult.

Due to the redesign, the victimization screening procedure used in 1995 and later years was meant to elicit a more complete tally of victimization incidents than the one used in 1989. For instance, it specifically asked whether respondents had been raped or otherwise sexually assaulted, whereas the 1989 screener did not. See *Effects of the Redesign on Victimization Estimates* (Kindermann, Lynch, and Cantor 1997) for more details on this issue.

In 2003, in accordance with changes to the Office of Management and Budget's standards for the classification of federal data on race and ethnicity, the NCVS item on race/ethnicity was modified. A question on Hispanic origin is now followed by a question on race. The new race question allows the respondent to choose more than one race and delineates Asian as a separate category from Native Hawaiian or Other Pacific Islander. An analysis conducted by the Demographic Surveys Division at the U.S. Census Bureau showed that the new race question had very little impact on the aggregate racial distribution of NCVS respondents, with one exception: there was a 2-percentage-point decrease in the percentage of respondents who reported themselves as White. Due to changes in race/ethnicity categories, comparisons of race/ethnicity across years should be made with caution.

In 2007, three changes were made to the NCVS for budgetary reasons. First, the sample was reduced by 14 percent beginning in July 2007. Second, to offset the impact of sample reduction, first-time interviews, which are not traditionally used in the production of the NCVS estimates, were included. Since respondents tend to report more victimization during first-time interviews than in subsequent interviews (in part, because new respondents tend to recall events having taken place at a time that was more recent than when they actually occurred), weighting adjustments were used to counteract a possible upward bias in the survey estimates. Using first-time interviews helped to ensure that the overall sample size would remain consistent with that in previous years. Lastly, in July 2007, the use of CATI as an interview technique was discontinued, and interviewing was conducted using only CAPI. For more details, see *Criminal Victimization, 2007* (U.S. Department of Justice 2008).

SCS design changes. The SCS questionnaire wording has been modified in several ways since its inception. Modifications have included changes in the series of questions pertaining to "fear" and "avoidance" between all survey years, beginning in 1995; changes in the definition of "at school" in 2001; changes in the introduction to, definition of, and placement of the item about "gangs" in 2001; and expansion of the single "bullying" question to include a series of questions in

2005 and including the topic of cyber-bullying in 2007. For more details, see *Student Victimization in U.S. Schools: Results From the 2005 School Crime Supplement to the National Crime Victimization Survey* (Bauer et al. 2008) and *Indicators of School Crime and Safety: 2008* (Dinkes, Kemp, and Baum 2009).

In addition, the reference time period for the 2007 SCS was revised from "the last 6 months" to "this school year." The change in reference period resulted in a change in eligibility criteria for participation in the 2007 SCS to include household members between ages 12 and 18 who had attended school at any time during the school year instead of during the 6 months preceding the interview, as in earlier surveys. This change was largely based on feedback obtained from students ages 12 to 18 during cognitive laboratory evaluations conducted by the U.S. Census Bureau. These respondents revealed they were not being strict in their interpretation of the 6-month reference period and were responding based on their experiences during the entire school year. Analyses of 2007 SCS data showed that estimates from 2007 are comparable to those from previous years. No change in reference period was made for criminal victimizations reported in the main NCVS.

Comparisons with related surveys. NCVS/SCS data have been analyzed and reported in conjunction with several other surveys on crime, safety, and risk behaviors. (See *Indicators of School Crime and Safety, 2008* [Dinkes, Kemp, and Baum 2009].) These include both NCES and non-NCES surveys. There are four NCES surveys: the School Safety and Discipline Questionnaire of the 1993 National Household Education Survey; the Teacher Questionnaire (specifically, the teacher victimization items) of the 1993–94, 1999–2000, 2003–04, and 2007–08 Schools and Staffing Survey; the Fast Response Survey System's Principal/School Disciplinarian Survey, conducted periodically; and the School Survey on Crime and Safety (SSOCS), conducted in 1999–2000, 2003–04, 2005–06, and 2007-08.

The non-NCES surveys and studies include the Youth Risk Behavior Surveillance System (YRBSS), a national and state-level epidemiological surveillance system developed by the Centers for Disease Control and Prevention (CDC) to monitor the prevalence of youth behaviors that most influence health; the School Associated Violent Death Study (SAVD), a study developed by the CDC (in conjunction with the U.S. Departments of Education and Justice) to describe the epidemiology of school-associated violent death in the United States and identify potential risk factors for these deaths; the Supplementary Homicide Reports (SHR), a part of the Uniform Crime Reporting (UCR) program conducted by the Federal Bureau of Investigation to provide incident-level information on criminal homicides; and the Web-based Injury Statistics Query and Reporting System Fatal (WISQARS Fatal), which provides data on injury-related mortality collected by the CDC.

Readers should exercise caution when doing cross-survey analyses using these data. While some of the data were collected from universe surveys, most were collected from sample surveys. Also, some questions may appear the same across surveys when, in fact, they were asked of different populations of students, in different years, at different locations, and about experiences that occurred within different periods of time. Because of these variations in collection procedures, timing, phrasing of questions, and so forth, the results from the different sources are not strictly comparable.

Contact Information

For content information on the SCS, contact:

NCES

Monica Hill
Phone: (202) 502-7379
E-mail: monica.hill@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

BJS

Michael Rand
Phone: (202) 616-3494
E-mail: randm@ojp.usdoj.gov

Methodology and Evaluation Reports

The reports listed below were either published by the U.S. Department of Education, National Center for Education Statistics (indicated by an NCES number), by the U.S. Department of Justice, Bureau of Justice Statistics, or were jointly published. See the technical notes in each report for a discussion of methodology.

General

U.S. Department of Justice, Bureau of Justice Statistics. (2008). *Criminal Victimization, 2007* (NCJ-224390). U.S. Department of Justice. Washington, DC: Bureau of Justice Statistics.

Uses of Data

Addington, L.A., Ruddy, S,A., Miller, A.K., and DeVoe, J.F. (2002). *Are America's Schools Safe? Students Speak Out: 1999 School Crime Supplement* (NCES 2002-331). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Bauer, L., Guerino, P., Nolle, K.L., and Tang, S. (2008). *Student Victimization in U.S. Schools: Results From the 2005 School Crime Supplement to the National Crime Victimization Survey* (NCES 2009-306). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chandler, K.A., Chapman, C.D., Rand, M.R., and Taylor, B.M. (1998). *Students' Reports of School Crime: 1989 and 1995* (NCES 98-241/NCJ-169607). National Center for Education Statistics, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

DeVoe, J.F., and Bauer, L. (2010). *Student Victimization in U.S. Schools: Results From the 2007 School Crime Supplement to the National Crime Victimization Survey* (NCES 2010-319). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

DeVoe, J.F., and Kaffenberger, S. (2005). *Student Reports of Bullying: Results From the 2001 School Crime Supplement to the National Crime Victimization Survey* (NCES 2005–310). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

DeVoe, J.F., Peter, K., Noonan, M., Snyder, T.D., and Baum, K. (2005). *Indicators of School Crime and Safety: 2005* (NCES 2006-001/NCJ-210697). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Dinkes, R., Cataldi, E.F., Kena, G., and Baum, K. (2006). *Indicators of School Crime and Safety: 2006* (NCES 2007-003/NCJ-214262). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Dinkes, R., Cataldi, E.F., and Lin-Kelly, W. (2007). *Indicators of School Crime and Safety: 2007* (NCES 2008-021/NCJ-219553). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Dinkes, R., Kemp, J., and Baum, K. (2009). *Indicators of School Crime and Safety: 2008* (NCES 2009-022/NCJ-226343). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Dinkes, R., Kemp, J., and Baum, K. (2009). *Indicators of School Crime and Safety: 2009* (NCES 2010–012/NCJ 228478). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Ruddy, S., Bauer, L., and Neiman, S. (2010). *A Profile of Criminal Incidents at School: Results From the 2003–05 National Crime Victimization Survey Crime Incident Report* (NCES 2010-318). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Survey Design

Kindermann, C., Lynch, J., and Cantor, D. (1997). *Effects of the Redesign on Victimization Estimates* (NCJ-164381). U.S. Department of Justice. Washington, DC: Bureau of Justice Statistics.

2. SCHOOL SURVEY ON CRIME AND SAFETY

Overview

he School Survey on Crime and Safety (SSOCS) collects extensive crime and safety data from principals and school administrators of public schools. The survey builds on an earlier survey on school crime and safety conducted in 1997 using the Fast Response Survey System (FRSS). SSOCS focuses on incidents of specific crimes and offenses and a variety of specific discipline issues in public schools. It also covers characteristics of school policies, school violence prevention programs and policies, and school characteristics that have been associated with school crime. The survey is conducted with nationally

representative samples of regular public primary, middle, high, and combined schools in the 50 states and the District of Columbia. The sample does not include special education, alternative, or vocational schools; schools in the U.S. outlying areas and Puerto Rico, overseas Department of Defense schools, newly closed schools, home schools, Bureau of Indian Education schools, nonregular schools, ungraded schools, and schools with a high grade of kindergarten or lower.

Purpose. To collect detailed information on crime and safety from the schools' perspective; and to provide estimates of school crime, discipline, disorder, programs, and policies.

Components. SSOCS consists of a single questionnaire that is completed by principals or the person most knowledgeable about crime and safety issues at the school. Sections of the SSOCS questionnaire are composed of items about specific topics, including school practices and programs, parent and community involvement at school, school security, staff training, limitations on crime prevention, frequency of crime and violence at school, number of incidents, disciplinary problems and actions, and school characteristics.

Periodicity. SSOCS is administered to public primary, middle, high, and combined school principals in the spring of even-numbered school years. SSOCS is administered at the end of the school year to allow principals to report the most complete information possible. SSOCS was first administered in the spring of the 1999–2000 school year (SSOCS:2000). It has since been administered in the spring of the 2003–04, 2005–06, 2007–08, and 2009–10 school years (SSOCS:2004, SSOCS:2006, SSOCS:2008, and SSOCS:2010). A sixth collection is planned for the 2011–12 school year.

Uses of Data

SSOCS provides school-level data on crime and safety on the frequency of violence, the nature of the school environment, and the characteristics of school violence prevention programs. Such national data are valuable to policymakers and researchers who need to know what policies and programs are in place, what the level of crime is and how it is changing, and what disciplinary actions schools are taking. Some of the topics that may be examined are the following:

- Frequency and types of crimes at schools, including homicide, rape, sexual battery, attacks with or without weapons, robbery, theft, and vandalism;
- Frequency and types of disciplinary actions such as expulsions, transfers, and suspensions for selected offenses;
- Perceptions of other disciplinary problems such as bullying, verbal abuse, and disorder in the classroom;
- School policies and programs concerning crime and safety; and
- Pervasiveness of student and teacher involvement in efforts that are intended to prevent or reduce school violence.

The survey data also support analyses of how these topics are related to each other and how they are related to various school characteristics.

Sample Design

A stratified sample design is used to select schools for SSOCS. The sampling frame for SSOCS is constructed from the NCES Common Core of Data (CCD) Public Elementary/Secondary School Universe data file. Only "regular" schools (i.e., excluding special education, alternative, or vocational schools; schools in other U.S. jurisdictions; and schools that teach only prekindergarten, kindergarten, or adult education) are eligible for SSOCS. A stratified sample of 3,370 public schools was selected for SSOCS:2000; 3,740 public schools for SSOCS:2004; 3,570 public schools for SSOCS:2006; 3,480 for SSOCS:2008; and 3,476 for SSOCS:2010.

The same general sample design is used for each SSOCS. For sample allocation purposes, strata are defined by instructional level, type of locale, and enrollment size. Black, Hispanic, and other race/ethnicity status, and region were used as sorting variables in the sample selection process for SSOCS:2000, SSOCS:2004, SSOCS:2006, and SSOCS:2008 to induce additional implicit stratification. Beginning with SSOCS:2010, percent White enrolment and region were used as sorting variables. The three explicit and two implicit stratification variables have been shown to be related to school crime and thus create meaningful strata for this survey. The sample is designed to provide reasonably precise cross-sectional estimates for selected subgroups of interest.

Although the same design was used to allocate the sample across strata for all administrations of SSOCS, the calculation of the total initial sample differed between SSOCS:2000 and later SSOCS administrations. Without the experience of prior

administrations, stratum response rates had to be estimated for SSOCS:2000 when determining the number of sample cases within each stratum. In contrast, later administrations took advantage of the lessons learned from the prior data collection and used the prior stratum response rates to determine the proper size of the initial sample.

Data Collection and Processing

The data collection phase consists of (1) a mailout/ mailback stage; and (2) a telephone follow-up stage.

Reference dates. Data for SSOCS are collected at the end of even-numbered school years to allow principals to report the most complete information possible. For example, data collected in 2000 pertain to the 1999–2000 school year.

Data collection. SSOCS is conducted as a mail survey with telephone follow-up. Advance letters and, in some cases, e-mails, are sent to sampled schools informing them that they have been selected for SSOCS and describing the survey. SSOCS questionnaires are mailed to administrators with a cover letter describing the importance of the survey and a brochure providing additional information about it.

Starting approximately 1-2 weeks after the first questionnaire mailing, follow-up telephone prompts are used to verify that the questionnaire was received and to encourage survey response. As an alternative to replying by mail, data are also accepted by fax submission and by telephone.

After the data collection ends, returned questionnaires are examined for quality and completeness using both manual and computerized edits. Key items are identified. Depending on the total number of items that have missing or problematic data, and on whether these items have been designated as key items, data quality issues are resolved by recontacting the respondents or by imputation.

Editing. The survey questionnaires are reviewed to match survey responses with the appropriate values to be entered. After the data are key-entered, they are run through a series of editing programs: first, to determine whether a returned questionnaire can be considered complete; subsequently, to check data for consistency, valid data value ranges, and skip patterns.

Weighting

Data are weighted to compensate for differential probabilities of selection and to adjust for the effects of nonresponse.

Sample weights allow inferences to be made about the population from which the sample units are drawn. Because of the complex nature of the SSOCS sample design, these weights are necessary to obtain population-based estimates, to minimize bias arising from differences between responding and nonresponding schools, and to calibrate the data to known population characteristics in a way that reduces sampling error.

An initial (base) weight is first determined within each stratum by calculating the ratio of the number of schools available in the sampling frame to the number of schools selected. Because some schools refuse to participate, the responding schools do not necessarily constitute a random sample of the schools in the stratum. In order to reduce the potential of bias from nonresponse, weighting classes are determined by using a statistical algorithm similar to CHAID (i.e., chi-square automatic interaction detector) to partition the sample such that schools within a weighting class are homogeneous with respect to their probability of responding. The predictor variables for the analysis are school instructional level; locale; region; enrollment size; percent enrollment of Black, Hispanic, and other race/ethnicity students (or percent White enrollment for SSOCS:2010 and beyond); student-to-teacher ratio; percentage of students eligible for free or reduced-price lunch; and number of full-time-equivalent teachers. When the number of responding schools in a class is small, the weighting class is combined with another to avoid the possibility of large weights. After combining the necessary classes, the base weights are adjusted to produce nonresponse- adjusted weights, so that the weighted distribution of the responding schools resembles the initial distribution of the total sample.

The nonresponse-adjusted weights are then poststratified to calibrate the sample to known population totals in order to reduce bias in the estimates due to undercoverage. Two-dimension margins are set up for the poststratification: (1) instructional level and school enrollment size; and (2) instructional level and locale. An iterative process, known as the raking ratio adjustment, brings the weights into agreement with the known control totals. To be effective, the variables that define the poststrata must be correlated with the outcome of interest (school crime, for example). All three variables—instructional level, school enrollment size, and locale—have been shown to be correlated with school crime (Miller 2004).

Imputation

Completed SSOCS surveys contain some level of item nonresponse after the conclusion of the data collection

phase. Imputation procedures were used to impute missing values of key items in SSOCS:2000 and missing values of all items in each subsequent SSOCS. All imputed values are flagged as such.

SSOCS:2000. In SSOCS:2000, only the key data items with missing data in the file were imputed. Depending on the type of data to be imputed and the extent of missing values, a number of techniques—including hot-deck imputation, hot-deck imputation with collapsed imputation cell, logical imputation, and mean imputation—were employed.

SSOCS:2004 and beyond. In subsequent collections, imputation procedures were used to create values for all questionnaire items with missing data. This procedural change from SSOCS:2000 was implemented because the analysis of incomplete datasets may cause different users to arrive at different conclusions, depending on how the missing data are treated. The imputation methods used in SSOCS:2004 and later surveys were tailored to the nature of each survey item. Four methods were used: aggregate proportions, logical, best match, and clerical.

Future Plans

NCES plans to conduct SSOCS every 2 years in order to provide continued updates on crime and safety in U.S. public schools. SSOCS will next be administered in the 2011–12 school year.

Sampling Error

The estimators of sampling variances for SSOCS statistics take the SSOCS complex sample design into account. Both replication and Taylor Series methods are used to estimate sampling errors in SSOCS.

SSOCS utilizes the jackknife replication method, which involves partitioning the entire sample into a set of groups (replicates) based on the actual sample design of the survey. Survey estimates can then be produced for each of the replicates by utilizing replicate weights that mimic the actual weighting procedures used in the full sample. The variation in the estimates computed for the replicates can then be used to estimate the sampling errors of the estimates for the full sample. A total of 50 replicate weights were defined for each SSOCS.

Another approach to the valid estimation of sampling errors for complex sample designs is to use a Taylor series approximation. To produce standard errors using a Taylor series program, two variables are required (to identify the stratum and the Primary Sampling Unit [PSU]). The stratum-level variable is the indicator of the variance estimation stratum from which the unit was selected. The PSU is an arbitrary numeric identification number for the unit within the stratum.

Unit Nonresponse

A response rate is the ratio of the number of completed questionnaires to the number of cases sampled and eligible to complete the survey. All of the response rates are weighted to account for different probabilities of selection. Schools that are determined to be ineligible to participate in the survey (e.g. special education, alternative, or vocational schools; schools in other U.S. jurisdictions; and schools that teach only prekindergarten, kindergarten, or adult education) are not included in the calculation of response rates. For SSOCS:2000, the weighted response rate was 70 percent and the final number of respondents was about 2,270. For SSOCS:2004, the weighted response rate was 77 percent and the final number of respondents was about 2,770. For SSOCS:2006, the weighted response rate was 81 percent and the final number of respondents was about 2,720. For SSOCS:2008, the weighted response rate was 77 percent and the final number of respondents was about 2,560. As of the date of this publication, response rates were not yet available for SSOCS:2010. (See table 21 for weighted unit response rates by selected characteristics.)

Nonresponse bias analyses were conducted to determine if substantial bias is introduced due to school nonresponse. In SSOCS:2000, a CHAID analysis was conducted to group table cells to efficiently adjust for nonresponse, and regression analysis was used to confirm the choice of variables that resulted from the CHAID analysis. The study found virtually no significant differences in the estimates when comparing the initial nonresponse adjustments and the additional adjustments that were adopted based on the CHAID analysis. This suggests that much of the variation in response rates was captured in the original sampling strata. The adjustments to the weights were retained, despite their small impact, based on theoretical considerations that suggest they should be effective in attenuating nonresponse biases for a broad range of statistics.

In the 2004, 2006, and 2008 SSOCS, a number of analyses compared nonresponding and responding schools. The base-weighted distributions of the eight sampling frame variables—instructional level; type oflocale; region; school enrollment size; percent Black, Hispanic, and other race/ethnicity enrollment; student-to-teacher ratio; percentage of students eligible for free or reduced-price lunch; and number of full-time-equivalent teachers—were compared for responding and nonresponding schools. Then the differences and the full sample, using the base sampling weight,

Table 21. Weighted unit response rates, by selected school characteristics: 2000, 2004, 2006, and 2008

School characteristics	2000	2004	2006	2008
Total	70.0	77.2	81.3	77.2
Instructional level				
Primary	69.0	76.5	83.0	77.0
Middle	69.7	75.5	79.9	77.0
High school	71.0	77.8	78.8	76.2
Combined	79.6	84.9	75.7	80.8
Enrollment size				
Less than 300	76.3	86.0	83.2	83.3
300–499	70.9	77.8	84.7	76.7
500–999	67.5	72.8	79.9	76.2
1,000 or more	61.1	71.1	72.5	68.6
Urbanicity[1]				
City	63.6	69.0	75.4	69.4
Suburb	67.5	72.5	80.3	73.1
Town	75.4	84.9	86.7	84.6
Rural	77.0	86.1	85.5	83.9
Percent Black, Hispanic, and other race/ethnicity				
Less than 5 percent/ missing[2]	77.8	85.9	89.5	84.3
5 to 19 percent	71.3	77.7	82.8	80.8
20 to 49 percent	65.4	75.8	79.3	76.7
50 percent or more	64.6	71.4	76.7	71.4
Region				
Northeast	64.1	71.7	78.0	69.5
Midwest	74.0	80.8	83.2	80.8
South	77.1	79.8	82.5	79.7
West	64.3	75.7	80.9	74.6

[1]Starting with SSOCS:2008, a 12-category urban-centric locale variable from the NCES Common Core of Data (CCD) file was used; it was collapsed into 4 categories: city, suburb, town, and rural. Prior SSOCS collections used an 8-category CCD variable collapsed into 4 categories: city, urban fringe, town, and rural. Therefore, caution should be exercised when making direct comparisons between the 2008 and prior SSOCS collections.

[2]Beginning in 2008, there was no missing data for the race/ethnicity variable. This variable was imputed prior to sampling.

SOURCE: Chaney, B., Chowdhury, S., Chu, A., Lee, J., and Wobus, P. (2004). *School Survey on Crime and Safety (SSOCS) 2000 Public-Use Data Files, User's Manual, and Detailed Data Documentation* (NCES 2004-306). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Guerino, P., Hurwitz, M.D., Kaffenberger, S.M., Hoaglin, D.C., and Burnaska, K. (2007). *2003–04 School Survey on Crime and Safety Data File User's Manual* (NCES 2007-335). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Neiman, S., and DeVoe, J.F. (2009). *Crime, Violence, Discipline, and Safety in U.S. Public Schools: Findings From the School Survey on Crime and Safety: 2007–08* (NCES 2009-326). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Nolle, K.L., Guerino, P., and Dinkes, R. (2007). *Crime, Violence, Discipline, and Safety in U.S. Public Schools: Findings From the School Survey on Crime and Safety: 2005–06* (NCES 2007-361). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

between the respondent sample, using the final weight, were examined with respect to all eight sampling frame variables. Generally, the differences were not significant, leading to the conclusion that nonresponse bias is not an issue.

Item Nonresponse

Generally, item response rates were quite high. Because a more extensive follow-up was conducted when nonresponse was present for key items, item response rates were often higher for key items than for other questionnaire items.

For the 2008 SSOCS, weighted item response rates for individual items within the questionnaire ranged from 72 to 100 percent.

Of the 241 subitems in the 2008 SSOCS questionnaire, only 13 had response rates below 85 percent, and a nonresponse bias analysis was conducted on these 13 items. The detected bias was not deemed problematic enough to suppress any items from the data file.

Contact Information

For content information on SSOCS, contact:

Monica Hill
Phone: (202) 502-7379
E-mail: monica.hill@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

Methodology and Evaluation Reports

General

Chaney, B., Chowdhury, S., Chu, A., Lee, J., and Wobus, P. (2003). *2000 School Survey on Crime and Safety: Detailed Data Documentation* (NCES 2004-307). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chaney, B., Chowdhury, S., Chu, A., Lee, J., and Wobus, P. (2003). *2000 School Survey on Crime and Safety: Public-Use Data File User's Manual* (NCES 2004-308). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chaney, B., Chowdhury, S., Chu, A., Lee, J., and Wobus, P. (2004). *School Survey on Crime and Safety (SSOCS) 2000 Public-Use Data Files, User's Manual, and Detailed Data Documentation* (NCES 2004-306). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Guerino, P., Hurwitz, M.D., Kaffenberger, S.M., Hoaglin, D.C., and Burnaska, K. (2007). *2003–04 School Survey on Crime and Safety Data File User's Manual* (NCES 2007-335). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Izrael, D., and deFriesse, F.A. (2007). *2003–04 School Survey on Crime and Safety Public-Use Data File Codebook* (NCES 2007-333). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ruddy, S.A., Neiman, S., Bauer, L., Swaim, N.L., Thomas, T.L., and Parmer, R.J. (2009). 2*005-06 School Survey on Crime and Safety (SSOCS) Survey Documentation for Data Users* (NCES 2010-320). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ruddy, S.A., Neiman, S., Hryczaniuk, C.A., Thomas, .L., and Parmer, R.J. (2010). *2007–08 School Survey on Crime and Safety (SSOCS) Survey Documentation for Public-Use Data File Users* (NCES 2010-307). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Tonsager, K., Neiman, S., Hryczaniuk, C., and Guan, X. (2010). *School Survey on Crime and Safety: 2007–08, Public-Use Data File Codebook* (NCES 2010-334). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wallace, E.C., Neiman, S., Foster, K.M., and Guan, X. (2009). 2*005–06 School Survey on Crime and Safety (SSOCS) Public-Use Data File Codebook* (NCES 2009-312). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Uses of Data

DeVoe, J.F., Peter, K., Noonan, M., Snyder, T.D., and Baum, K. (2005). *Indicators of School Crime and Safety: 2005* (NCES 2006-001/NCJ-210697). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice

Programs, U.S. Department of Justice. Washington, DC.

Dinkes, R., Cataldi, E.F., Kena, G., and Baum, K. (2006). *Indicators of School Crime and Safety: 2006* (NCES 2007-003/NCJ 214262). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Dinkes, R., Cataldi, E.F., and Lin-Kelly, W. (2007). *Indicators of School Crime and Safety: 2007* (NCES 2008-021/NCJ-219553). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Dinkes, R., Kemp, J., and Baum, K. (2009). *Indicators of School Crime and Safety: 2008* (NCES 2009–022/NCJ 226343). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Guerino, P., Hurwitz, M.D., Noonan, M.E., and Kaffenberger, S.M. (2006). *Crime, Violence, Discipline, and Safety in U.S. Public Schools: Findings From the School Survey on Crime and Safety: 2003–04* (NCES 2007-302REV). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Miller, A.K. (2004). *Violence in U.S. Public Schools: 2000 School Survey on Crime and Safety* (NCES 2004-314REV). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Neiman, S., and DeVoe, J.F. (2009). *Crime, Violence, Discipline, and Safety in U.S. Public Schools: Findings From the School Survey on Crime and Safety: 2007–08* (NCES 2009-326). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Nolle, K.L., Guerino, P., and Dinkes, R. (2007). *Crime, Violence, Discipline, and Safety in U.S. Public Schools: Findings From the School Survey on Crime and Safety: 2005–06* (NCES 2007-361). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 29: High School Transcript (HST) Studies

The value of school transcripts as objective, reliable measures of crucial aspects of students' educational experiences is widely recognized. NCES high school transcript studies collect information that is contained on the student high school record—i.e., courses taken while attending secondary school; information on credits earned; year and term a specific course was taken; and, final grades. When available, information on class rank and standardized scores is also collected. Once collected, information (e.g., course name, credits earned, course grades) is transcribed and standardized (e.g., credits and credit hours standardized to a common metric) and can be linked back to the student's questionnaire or assessment data.

Transcripts include information that is considered to be the official and fixed record regarding student course taking behaviors. This information is considered to be more accurate than student self-report information and represents a record of courses taken by the student. This information can be used to examine course-taking patterns of students and to predict future education outcomes.

Since 1982, NCES has conducted nine high school transcript studies: six high school transcript studies (HSTS) associated with the National Assessment of Educational Progress (NAEP), and three high school transcript studies as part of the Longitudinal Studies Program. Some of the key terms related to high school transcript studies are defined below.

Advanced Placement (AP). The AP Program is designed to prepare students to take the advanced placement examinations given by the Educational Testing Service (ETS). Students who pass these tests may be given credit and/or be exempted from requirements in colleges and universities based on their scores. Colleges and universities make their own rules regarding what tests to accept and the scores needed for credit or exemptions.

Carnegie unit. A factor used to standardize all credits indicated on transcripts across the study. The Carnegie unit is a strictly time-based reference for measuring secondary school attainment used by American universities and colleges. A single Carnegie unit is equal to 120 hours of classroom time over the course of a year at the secondary American high school level. Strictly speaking, this breaks down into a single 1-hour meeting on each of 5 days per week for a total of 24 weeks per year. However, knowing that classes usually meet for 50 minutes yields a value of 30 weeks per year. A semester (one-half of a full year) earns 1/2 Carnegie unit.

Catalog. A document compiled by a school or a district listing all available courses that are offered by the school and a description of those courses. Curriculum specialists review catalogs and use them to determine the appropriate Classification of Secondary School Courses (CSSC) code for each course.

Classification of Secondary School Courses (CSSC). A coding system employed for the purpose of standardizing high school transcript transcripts. The CSSC is a modification of the Classification of Instructional Programs (CIP) code used for

NAEP HIGH SCHOOL TRANSCRIPT STUDIES:

- National Assessment of Educational Progress High School Transcript Studies

LONGITUDINAL SURVEYS HIGH SCHOOL TRANSCRIPT STUDIES:

- High School and Beyond Longitudinal Study High School Transcript Study
- National Education Longitudinal Study of 1988 High School Transcript Study
- Education Longitudinal Study of 2002 High School Transcript Study

classifying college courses and contains approximate 2,300 course codes. Each CSSC code contains six digits. The first two digits identify the main program area, the second two digits represent a subcategory of courses within the main program area, and the final two digits define the specific course. For example, for the CSSC code 400522, the first two digits (40) define the Physical Sciences program area, the middle two digits (05) define the Chemistry subcategory, and the final two digits (22) define the course Advanced Chemistry.

Course offerings file. An high school transcript study data file that provides a comprehensive list of the courses offered in the schools included in the study. A CSSC code is associated with each course title.

International Baccalaureate (IB). A nonprofit educational foundation program consisting of a comprehensive 2-year international curriculum that allows students to fulfill the requirements of their national or state education systems.

Secondary School Taxonomy. The framework initially used by high school transcript for analyzing transcript data. The taxonomy divides high school coursework into three distinct curricula: academic, vocational, and personal/other.

Taxonomy. The classification of items into larger categories. In high school transcript studies, the items are specific secondary school courses (e.g., composition, first-year algebra, AP biology, American government) that are classified into course subject categories, as organized according to the Secondary School Taxonomy (SST), which is based on course content and level.

Tests and Honors file. A data file providing a list of honors and standardized test results, including SAT and ACT scores, that are found in the transcripts.

Transcript. A student's secondary school record containing courses taken, grades, graduation status, and attendance. In addition, it often includes scores from assessments, such as the PSAT, SAT, ACT, and a list of honors.

Transcript file. A data file providing a complete list of all courses appearing in the transcripts of students sampled in the study.

1. NAEP High School Transcript Studies

Since 1982, NCES has conducted six high school transcript studies (HSTS) associated with the National Assessment of Educational Progress (NAEP). NAEP has collected transcript data in 1987, 1990, 1994, 1998, 2000, 2005 and 2009 (see chapter 18). The results for 2009 HSTS will be reported in winter of 2011. Since information on 2009 HSTS is not yet available, 2009 HSTS will not be included in some sections of this chapter.

Components
Conducted in conjunction with NAEP, the 2009, 2005, 2000, 1998, 1994, 1990 and 1987 HSTS collected information on course offerings and coursetaking patterns in the nation's schools. Transcript data can be used to show coursetaking patterns across years that may be associated with proficiency in subjects assessed by NAEP.

Transcripts were collected for twelfth-grade students who graduated high school by the end of the collection period. Most students also participated in the NAEP assessments earlier that same year. Specifically, the students included in the 2005 and 2000 HSTS participated in the NAEP twelfth-grade mathematics and science assessments in 2005 and 2000 respectively; the students included in the 1998 HSTS participated in the civics, reading, and writing assessments in 1998; the students included in the 1994 HSTS participated in the geography, reading, and U.S. history assessments in 1994; the students included in the 1990 HSTS participated in the mathematics, science, and reading assessments in 1990; the students included in the 1987 HSTS participated in the 1986 long-term trend NAEP assessments in mathematics and science.

Periodicity
High school transcript studies have been conducted by NCES in conjunction with the NAEP since 1982. NAEP has collected transcript data in 1987, 1990, 1994, 1998, 2000, 2005, and 2009.

Survey Design

Target Population
The target population for high school transcript studies conducted as part of longitudinal surveys included all students in public and private schools who participated in previous data collections. For example, the target population for the 2004 high school transcript study included students who been in-school sophomores in the 2001–02 school year, participated in both the base-year and first follow-up interviews, completed the

mathematics assessment in the base-year and first follow-up interviews, and had complete transcript information for the 2002–03 and 2003–04 school years. The 2004 high school transcript study included 14,710 of the originally selected sample members of ELS:2002 sophomores in the spring of 2002 who were respondents in both the base-year and first follow-up interviews.

Sample Design

The NAEP High School Transcript Studies were conducted using nearly identical methodologies and techniques. They include the 2005, 2000, 1998, 1994, 1990, and 1987 transcript studies.

The 2005 High School Transcript Study. The sample design for the 2005 HSTS was designed to achieve a nationally representative sample of public and private high school graduates in the class of 2005. For public schools, the HSTS sample was the 12th-grade public school sample for the 2005 NAEP mathematics and science assessments; that is, the HSTS sample included every eligible sampled 2005 NAEP 12th-grade public school that was contacted for the HSTS, whether or not they actually participated in the NAEP assessments. For private schools, the HSTS sample was a subsample from the 2005 NAEP 12th-grade private school sample for the mathematics and science assessments. This subsampling process was carried out because private schools were oversampled in the 2005 NAEP. For the HSTS, the sample design called for the private schools' sample size to be proportionate to their share of eligible students. Over 26,000 transcripts from graduates were collected for the 2005 HSTS from a sample of about 640 public schools and 80 private schools.

For NAEP-participating schools, only those that participated in the main NAEP mathematics and science assessments were eligible for the HSTS. Within these schools, the HSTS used the same NAEP mathematics and science student samples. For schools that were selected for NAEP but did not participate, graduates were randomly selected. Approximately 94 percent of the HSTS sampled students were enrolled in schools that also participated in the NAEP assessments. Around 63 percent of the participating HSTS students also participated in NAEP.

The 2000 High School Transcript Study. The 2000 HSTS school sample comprised all 320 12th-grade public schools and a subsample of the 620 12th-grade private schools selected for the 2000 NAEP. The objective of private school subsampling was to reverse the oversampling of private schools in the 2000 NAEP so that the private school students in the 2000 HSTS would be represented in proportion to their prevalence in the general 12th-grade student population. While in NAEP 2000, private schools were oversampled to meet explicit target sample sizes for reporting group in order to provide reliable NAEP estimates for such students; in HSTS 2000, however, the oversampling of private schools was reversed so that the private school students in HSTS were represented in proportion to their prevalence in the general 12th-grade student population.

Because sampling was performed in most high schools prior to graduation, not all sampled students were, in fact, graduates. Only graduates, however, were eligible for the transcript study. From the exit status of the students, it was determined that of the 23,440 students in the sample, 21,090 actually graduated by October 2000 and 2,360 did not. From the 21,090 graduates, 20,930 transcripts were collected and processed. That is, 99 percent of the transcripts of eligible students were obtained.

The 1998 High School Transcript Study. The 1998 HSTS sample is nationally representative at both the school and student levels. The sample was composed of schools selected for the NAEP main sample that had 12th-grade classes and were within the 58 primary sampling units (PSUs) selected for the HSTS study. A subsample of 320 schools was selected from the eligible NAEP sample, consisting of 270 public schools and 50 nonpublic schools. In order to maintain as many links as possible with NAEP scores, replacement schools that were used in NAEP were also asked to participate in the transcript study, as opposed to sampling the NAEP refusal schools. Of the 320 schools in the original sample, 260 participated, of which 230 cooperated with both NAEP and HSTS and maintained links between students' transcript and NAEP data.

A total of 28,760 students were selected for inclusion in the HSTS study. Of these, 27,180 students were from schools that maintained their NAEP administration schedules and were identified by their NAEP booklet numbers. Another 500 students were from schools that participated in NAEP but had lost the link between student names and NAEP booklet numbers, and 1,080 were from schools that did not participate in NAEP. Of the 28,760 students in the original sample, 25,250 were deemed eligible for the transcript study, and 24,220 transcripts were collected and processed.

The 1994 High School Transcript Study. The 1994 HSTS sample of schools was nationally representative of all high schools in the United States. A subsample of 330 public schools and 50 private schools was

drawn from the lists of eligible NAEP public and private schools. One of these schools had no 12th-grade students and was not included in the HSTS study. Of the 380 remaining schools, 340 participated in the 1994 HSTS. The student sample was representative of graduating seniors from each school. Only those students were included whose transcripts indicated that they had graduated between January 1, 1994, and November 21, 1994. Approximately 90 percent of students in the 1994 HSTS also participated in the 1994 NAEP. The remaining students were sampled specifically for the transcript study, either because their schools did not agree to participate in the 1994 NAEP or because the schools participated in NAEP but did not retain the lists linking NAEP IDs to student names. The 1994 HSTS also included special education students who were excluded from the 1994 NAEP. High school transcripts were collected for 25,500 students from an eligible sample of 26,050 students.

The 1990 High School Transcript Study. The sample of schools was nationally representative of schools with a grade 12 or having 17-year-old students. (Some 380 schools were selected for the sample; some of these had no 12th-grade students.) The sample of students was representative of graduating seniors from each school. These students attended 330 schools that had previously been sampled for the 1990 NAEP. Approximately three-fourths of the sampled students had participated in the 1990 NAEP assessments. The remaining students attended schools that did not participate in NAEP or did not retain the lists linking student names to NAEP IDs. As with the later HSTS, only schools with a 12th grade were included, and only students who graduated from high school in 1990 were included. The 1990 HSTS also included special education students who had been excluded from the 1990 NAEP. In spring 1991, transcripts were requested for 23,270 students who graduated from high school in 1990; 21,610 transcripts were received.

The 1987 High School Transcript Study. The 1987 HSTS was conducted in conjunction with the long-term trend NAEP assessment. The schools in the 1987 HSTS were a nationally representative sample of 500 secondary schools that had been selected for the 1986 long-term trend NAEP assessments. The 1987 HSTS student sample represented an augmented sample of 1986 NAEP participants who were enrolled in the 11th grade and/or were 17 years old in the 1985–86 school year and who successfully completed their graduation requirements prior to fall 1987. The HSTS study included (1) students who were selected and retained for the 1986 NAEP assessment; (2) students who were sampled for the 1986 NAEP but were deliberately excluded due to severe mental, physical, or linguistic barriers; and (3) all students with disabilities attending schools selected for the 1986 assessment. Four of the participating schools had no eligible students without disabilities. Of the 500 schools selected for the HSTS study, 430 participated. There were 35,180 graduates in the sample, for whom 34,140 transcripts were received.

Data Collection and Processing

Data collection. The data collection procedures of the 2005 HSTS are discussed to illustrate the process. NAEP field workers requested sample materials for the 2005 HSTS when they first went to a school as part of the 2005 NAEP, and they collected these materials when they returned to the school for sampling. The sample materials included a list of courses offered for each of 4 consecutive years from school year 2001–02 through 2004–05; a completed School Information Form (SIF); and three sample transcripts, one representing a student taking “regular” courses, one with honors courses, and one with special education courses. For those students who were selected to participate in NAEP but who were classified as either having disabilities (SD) or English language learners (ELL), an SD/LEP questionnaire was completed for these students by the person most knowledgeable about the student. A School Questionnaire—which asked for information about school, teacher, and home factors that might relate to student achievement—was completed by a school official (usually the principal) as part of NAEP.

The SIF collected information about the school in general, sources of information within the school, course description materials, graduation requirements, grading practices, and the format of the school transcripts as part of the HSTS data collection process for non-NAEP participating schools.

In schools that did not participate in NAEP, the field worker first selected a sample of students, then requested transcripts for those students and followed the procedures for NAEP participants for reviewing and shipping transcripts. The SIF was also completed, and course catalogs for the past 4 school years were collected. The information in the catalogs was documented by completing the Course Catalog Checklist. At this point, the procedure was different from the one used for schools that participated in NAEP. Rather than obtaining and annotating three example transcripts, the field worker used the Transcript Format Checklist to annotate three actual transcripts from among those that were collected.

In the non-NAEP participating schools, the process of generating a sample of students began when the school

produced a listing of all students who graduated from the 12th grade during the spring or summer of 2005. This list was requested during the preliminary call placed to the school when it was determined that the school would participate in the HSTS. The following information was collected for each student in the HSTS: exit status; sex; date of birth (month/year); race/ethnicity; whether the student had a disability; whether the student was classified as limited English proficient; whether the student was receiving Title I services; and whether the student was a participant in the National School Lunch Program. These data were collected either with the list of 2005 graduates or after sampling, depending on which procedure was easier for the school.

Data processing. Each of the courses entered on the transcripts were coded using the Classification of Secondary School Courses (CSSC). For all NAEP transcript studies, courses appearing on student transcripts were coded to indicate whether they were transfer courses, held off campus, honors or above grade-level, remedial or below grade-level, or designed for students with limited English proficiency and/or taught in a language other than English.

Credit and grade information reported on transcripts also needed to be standardized. Standardization of credit information was based on the Carnegie unit, defined as the number of credits a student received for a course taken every day, one period per day, for a full school year.

The Computer-Assisted Coding and Editing (CACE) system was designed specifically for coding high school catalogs. CACE has two major components: (1) a component for selecting and entering the most appropriate CSSC code and "flags" for each course in a catalog; and (2) a component for matching each entry appearing on a transcript with the appropriate course title in the corresponding school's list of course offerings.

Each stage of the data coding and entering process included measures to ensure the quality and consistency of data. Measures to maintain the quality of data entry on transcripts included 100 percent verification of data entry; review of all transcripts where the number of credits reported for a given year (or the total number of credits) was not indicative of the school's normal course load or graduation requirements; and reconciliation of transcript IDs with the list of HSTS-valid IDs. Catalog coding reliability was maintained by conducting reliability checks. At least 10 percent of each school's course offerings were reentered by an experienced coder and the results compared with those of the original coder. If less than 90 percent of the entries agreed, the catalog was completely reviewed and any necessary changes were made. Agreement of 90 percent or better was found for approximately 85 percent of the school catalogs during the first review.

An additional quality check took place when the CACE files for a school were converted to delivery format. Reports listing frequencies of occurrences that might indicate errors were sent to the curriculum specialist for review. Each file was then assigned a status of 1 for complete, 2 for errors in transcript entry, 3 for errors in catalog coding and associations, or 4 for computer errors. A file with a status of 2, 3, or 4 was returned to Computer-Assisted Data Entry (CADE) and CACE for correction, a new report was generated, and the report was again reviewed. This process was repeated until the file had a status of 1, indicating that it was complete and correct.

Estimation Methods

Weighting. The weighting procedures are similar across the HSTS studies associated with NAEP. Only the 2005 NAEP HSTS procedures are described below. (For details on weighting in the other NAEP HSTS studies, see the relevant technical manuals.)

Two types of weights were created in the 2005 HSTS:

- HSTS base weights for all students who participated in the 2005 HSTS—that is, for whom a transcript was received and coded; and
- HSTS-NAEP linked weights for students who participated in both the 2005 HSTS and the 2005 NAEP. Linked weights were computed separately for mathematics and science assessment students. Each assessment sample represents the full population, so each of the two sets of assessment-linked weights aggregate separately to the population totals.

In each set of weights, the final weight attached to an individual student record reflected two major aspects of the sample design and the population surveyed. The first component, the base weight, reflected the probability of selection in the sample (the product of the probability of selecting the PSU, the probability of selecting the school within the PSU, and the probability of selecting the student within the school). The second component resulted from the adjustment of the base weight to account for nonresponse within the sample and to ensure that the resulting survey estimates of certain characteristics (race/ethnicity, size of

community, and region) conformed to those known reliably from external sources.

The final HSTS student weights were constructed in five steps:

(1) The student base weights (or design unbiased weight) were constructed as the reciprocal of the overall probability of selection.

(2) School nonresponse factors were computed, adjusting for schools that did not participate in the HSTS study. For the linked weights, adjustment factors were assigned for each session type (writing/civics, reading, and civics trend). The school nonresponse factors for the linked weights were also slightly different from the corresponding HSTS student weight school nonresponse factors to account for schools that refused to participate in NAEP.

(3) Student nonresponse factors were computed, adjusting the weights of responding students to account for nonresponding students. Definitions of responding and nonresponding students differed for the HSTS weights and the linked weights.

(4) Student trimming factors were generated to reduce the mean squared error of the resulting estimates. Another purpose of the trimming was to protect against a small number of large weights from dominating the resulting estimates of small domains of interest.

(5) The final step was poststratification, the process of adjusting weights proportionally so that they aggregate within certain subpopulations to independent estimates of these subpopulation totals. These independent estimates were obtained from Current Population Survey (CPS) estimates for various student subgroups. As the CPS estimates are associated with smaller sampling errors, this adjustment should improve the quality of the weights.

The linked student weights were constructed in a parallel manner, with some differences (e.g., the student base weight incorporated a factor for assignment to NAEP assessments). The school nonresponse factors were also slightly different for the linked weights to account for schools that refused to participate in the NAEP assessments. In addition, an extra nonresponse factor was computed for the linked weights to adjust for students whose transcripts were included in the HSTS study but who were absent from (or refused to participate in) a NAEP assessment. The trimming and poststratification steps for the linked weights were similar to those for the HSTS weights, with some differences. The missing transcript adjustments for the linked weights were very similar to those computed for HSTS weights.

Imputation. Imputation was done for missing data in the 1994, 1998, 2000, and 2005 High School Transcript Studies conducted in conjunction with NAEP.
In the 1994, 1998, 2000, and 2005 HSTS, it was not possible to obtain a transcript for a small percentage of high school graduates. In addition, some transcripts were considered unusable, since the number of standardized credits shown on the transcript was less than the number of credits required to graduate by the school. An adjustment is necessary in the weights of high school graduates with transcripts to account for missing and unusable transcripts. To do this adjustment correctly, it is necessary to have the complete set of high school graduates, with or without transcripts. Students who did not graduate were not included in this adjustment, but they were retained in the process for poststratification. There are a few students, however, for whom no transcripts were received and whose graduation status was unknown. Among these students, a certain percentage was imputed as graduating, based on the overall percentages of high school graduates. The remaining students were imputed as nongraduating. The imputation process was a standard (random within class) hot-deck imputation. For each student with unknown graduation status, a “donor” was randomly selected (without replacement) from the set of all students with known graduation status from the same region, school type, race/ethnicity, age class, school, and sex, in hierarchical order. The two race/ethnicity categories were (1) White, Asian, or Pacific Islander; and (2) Black, Hispanic, American Indian, or other. There were two age classes (born before 10/79; born during or after 10/79). Each student with known graduation status in a cell could be used up to three times as a donor for a student in the same cell with unknown graduation status. If insufficient donors were available within the cell, donors were randomly selected from students in another cell with similar characteristics to the cell in question. At the least, a donor had to be from the same region, type of school, race category, and age category.

Data Quality and Comparability

Sampling Error

Because of the HSTS multistage design, jackknife repeated replication was used for variance estimation in transcript studies associated with NAEP.

Table 22. Unweighted response rates for all eligible High School Transcript Study schools and students in each study: Various years, 1987-2005

Year	School response rate	Student coverage rate
2005	82	84[1]
2000	81	99
1998	88	98
1994	90	98
1990	87	93
1987[2]	87	97

[1] Weighted response rate.
[2] The 1987 HSTS was conducted in conjunction with the long-term trend NAEP assessment.
SOURCE: *The 1990 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1990, 1987, and 1982 High School Graduates* (No. ED360375). ERIC Document Reproduction Service. Washington, DC. Legum, S., Caldwell, N., Davis B., Haynes, J., Hill, T.J., Litavecz, S., Rizzo, L., Rust, K., Vo, N., and Gorman, S. (1997). *The 1994 High School Transcript Study Technical Report (*NCES 97-262). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Roey, S., Caldwell, N., Rust, K., Hicks, L., Lee, J., Perkins, R., Blumstein, E., and Brown, J. (2005). *The 2000 High School Transcript Study User's Guide and Technical Report* (NCES 2005-483). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Shettle, C., Cubell, M., Hoover, K., Kastberg, D., Legum, S., Lyons, M., Perkins, R., Rizzo, L., Roey, S., and Sickles, D. (2005). *The 2005 High School Transcript Study: The 2005 High School Transcript Study User's Guide and Technical Report* (NCES 2009-480). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Thorne, J. (1989). *High School Transcript Study, 1987* (No. ED315450). ERIC Document Reproduction Service. Washington, DC.

In the 2005 HSTS, a set of 62 replicate weights was attached to each record, one for each replicate. Variance estimation was performed by repeating the estimate procedure 63 times, once using the original full set of sample weights and once each for the set of 62 replicate weights. The variability among replicate estimates was used to derive an approximately unbiased estimate of the sampling variance. This procedure was used to obtain sampling errors for a large number of variables for the whole population and for specified subgroups.

Nonsampling Error

Coverage error. As the transcript studies associated with NAEP attempted to collect high school transcripts for all students selected for the assessment, whether or not they participated, transcripts for these students are included in the transcript study. Students who did not meet the graduation requirements established were excluded. Students with special education diplomas, certificates of attendance, and certificates of completion were also excluded, as were students with zero English credits and students with fewer than 16 Carnegie units. Because the NAEP studies collected data on the characteristics of excluded students, undercoverage bias can be quantified. Also, these studies were more inclusive in their transcript components than in their test or questionnaire administration. (See the section of "Sample Design".) It is believed that NAEP transcript studies had no transcript undercoverage due to exclusion of certain students.

Nonresponse error.

Unit nonresponse. There is unit nonresponse at both the school and student levels in HSTS. Response rates are presented in table 22.

An unweighted 82 percent of schools participated in the 2005 NAEP transcript study, higher than the 81 percent in the 2000 HSTS, but lower than the participation rate in the other NAEP transcript studies. Response rates varied with the characteristics of the sample school. For example, in 2005, despite a moderate overall response rate, only 57 percent of nonpublic schools responded.

At the student level, transcripts were obtained for 84 percent of eligible students in the 2005 HSTS (weighted), which is lower than the student-level response rate in the other transcript studies conducted in conjunction with NAEP. The response rate in the 2000 HSTS, 99 percent (unweighted), was the highest achieved in all six transcript studies.

Data Comparability

*Comparability of target populations***.** The target population of the 1987 NAEP HSTS has special features that affect its comparability to that of target populations in the other HSTS studies. The 1987 sample originated in a within-school representative sample of the schools' juniors/17-year-olds (students

born between October 1, 1968, and September 30, 1969). However, subsequent transfers into the school were given no chance of selection into the study; this fact qualifies how representative the within-school sample is and leaves it close to, but not precisely, a sample of the high school graduating class of 1987.

The 1990 HSTS sample originated within the 1990 NAEP sample of seniors/17-year-olds, but is further restricted to the seniors who in fact graduated in calendar year 1990. As such, it provides a nationally representative sample of 1990 high school graduates. Subsequent NAEP transcript collections have adhered to sample definitions that identify an unequivocally representative sample of graduating seniors.

*Sample inclusion and exclusion***.** A second issue concerns student sample inclusion and exclusion, especially with respect to students with disabilities and English language learners. The NAEP assessments collected information from school records about special education students. In the 1987 HSTS, the sample included students who were sampled for the assessment but deliberately excluded from it, as well as students with disabilities attending schools selected for the assessment. Thus, transcripts were collected for students excluded from the NAEP test as well as from the test-eligible sample. NAEP has carefully documented excluded students and identifies those who received testing accommodations.

Contact information

For content information about the High School Transcript Studies conducted in conjunction with NAEP, contact

Janis Brown
Phone: (202) 502-7482
E-mail: janis.brown@nces.ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

Methodology and evaluation reports

Alt, M.N., and Bradby, D. (1999). *Procedures Guide for Transcript Studies* (NCES Working Paper 1999-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Bradby, D., and Hoachlander, G. (1999). *1998 Revision of the Secondary School Taxonomy* (NCES Working Paper 1999-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Thorne, J. (1989). *High School Transcript Study, 1987* (No. ED315450). ERIC Document Reproduction Service. Washington, DC.

Caldwell, N. (1993). *The 1990 High School Transcript Study Data File User's Manual* (No. ED361354). ERIC Document Reproduction Service. Washington, DC.

Legum, S. (1993). *The 1990 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1990, 1987, and 1982 High School Graduates* (No. ED360375). ERIC Document Reproduction Service. Washington, DC.

Legum, S., Caldwell, N., Davis, B., Haynes, J., Hill, T.J., Litavecz, S., Rizzo, L., Rust, K., Vo, N., and Gorman, S. (1997). *The 1994 High School Transcript Study Technical Report (*NCES 97-262). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Perkins, R., Kleiner, B., Roey, S., and Brown, J. (2004). *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990-2000* (NCES 2004-455). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Roey, S., Caldwell, N., Rust, K., Blumstein, E., Krenzke, T., Legum, S., Kuhn, J., Waksberg, M., Haynes, J., and Brown, J. (2001). *The 1998 High School Transcript Study User's Guide and Technical Report* (NCES 2001-477). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Roey, S., Caldwell, N., Rust, K., Blumstein, E., Krenzke, T., Legum, S., Kuhn, J., Waksberg, M., Haynes, J., and Brown, J. (2001). *The 1998 High School Transcript Tabulations: Comparative Data on Credits Earned and Demographics for 1998, 1994, 1990, 1987, and 1982 High School Graduates* (NCES 2001-498). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Roey, S., Caldwell, N., Rust, K., Hicks, L., Lee, J., Perkins, R., Blumstein, E., and Brown, J. (2005). *The 2000 High School Transcript Study User's Guide and Technical Report* (NCES 2005-483). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Shettle, C., Cubell, M., Hoover, K., Kastberg, D., Legum, S., Lyons, M., Perkins, R., Rizzo, L., Roey, S., and Sickles, D. (2008) *The 2005 High School Transcript Study: The 2005 High School Transcript Study User's Guide and Technical Report.* (NCES 2009-480). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Shettle, C., Roey, S., Mordica, J., Perkins, R., Nord, C., Teodorovic, J., Brown, J., Lyons, M., Averett, C., and Kastberg, D. (2007). *America's High School Graduates: Results From the 2005 NAEP High School Transcript Study* (NCES 2007-467). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

2. Longitudinal Surveys High School Transcript Studies

Overview

Since 1982, NCES has conducted three high school transcript studies as part of the Longitudinal Studies Program. The first NCES-sponsored transcript study was conducted in 1982, as part of the first follow-up to the High School and Beyond (HS&B) Longitudinal Study (see chapter 7). In 1992, another transcript study was conducted in conjunction with the second follow-up to the National Education Longitudinal Study of 1988 (NELS:88) (see chapter 8). A third transcript study associated with the longitudinal study series was conducted in 2004, as part of the first follow-up to the Education Longitudinal Study of 2002 (ELS:2002) (see chapter 9).

Components

The 2004 High School Transcript Study. The ELS:2002 high school transcript data collection sought information about coursetaking from the student's official high school record—including courses taken while attending secondary school, information on credits earned, year and term a specific course was taken, and final grades. When available, other information was collected, including dates enrolled, reason for leaving school, and standardized test scores. Once collected, the information was transcribed and can be linked back to the student's questionnaire or assessment data collected by ELS. Due to the size and complexity of the file, and because of reporting variations by school, additional variables were constructed from the raw transcript file. These composite variables include standardized grade point average (GPA), high school academic program, total credits earned by subject, and others.

The 1992 High School Transcript Study. The NELS:1992 high school transcript data include detailed information about the types of degree programs, periods of enrollment, majors or fields of study, specific courses taken, grades and credits attained, and credentials earned.

The 1982 High School Transcript Study. The HS&B transcript data collection allows the study of the coursetaking behavior of the members of the 1980 sophomore cohort throughout their 4 years of high school. Data include a six-digit course number for each course taken; course credit, expressed in Carnegie units (a standard of measurement that represents one credit for the completion of a 1-year course); course grade; year course was taken; GPA; days absent; and standardized test scores.

Periodicity

High school transcript studies have been conducted by NCES as part of the Longitudinal Studies Program since 1982. Transcript studies associated with the Longitudinal Studies Program were conducted in 1982, 1992, and 2004.

Survey Design

Target Population

The target population for high school transcript studies conducted as part of longitudinal surveys included all students in public and private schools who participated in previous data collections. For example, the target population for the 2004 high school transcript study included students who been in-school sophomores in the 2001–02 school year, participated in both the base-year and first follow-up interviews, completed the mathematics assessment in the base-year and first follow-up interviews, and had complete transcript information for the 2002–03 and 2003–04 school years. The 2004 high school transcript study included 14,710 of the originally selected sample members of ELS:2002 sophomores in the spring of 2002 who were respondents in both the base-year and first follow-up interviews.

Sample Design

Sample design is essentially similar across the various administrations of the high school transcript studies: multistage, stratified, and clustered design.

The 2004 ELS High School Transcript Study. This study was conducted as part of the ELS:2002 first follow-up in 2004 (see chapter 9). A total of 1,550 out of 1,950 schools participated in the request for

transcripts for an unweighted participation rate of 79 percent. The base-year school weighted response rate is 95 percent. The course offerings response rate for base-year schools is 88 percent. Ninety-one percent(91 percent, weighted) of the entire student sample have some transcript information (14,920 out of 16,370 students).

Transcripts were collected from the school that the students were originally sampled from in the base year (which was the only school for most sample members) and from their last school of attendance if it was learned during the first follow-up student data collection that they had transferred. Incomplete records were obtained for sample members who had dropped out of school, had fallen behind the modal progression sequence, or were enrolled in a special education program requiring or allowing more than 12 years of schooling. For freshened students, transcripts were only collected from their senior year school. Transcripts were collected for regular graduates, dropouts, early graduates, and students who were homeschooled after their sophomore year.

The 1992 High School Transcript Study. This transcript study was conducted as part of the NELS:88 second follow-up (see chapter 8). A total of 2,260 schools were identified as longitudinal cohort eligible for the high school transcript study, in the second follow-up tracing of the NELS:88 first follow-up sample. Since the high school transcript study conducted as part of NELS:88 was limited to 1,500 schools for the full range of data (student, parent, teacher, school administrator, and transcript data) collection, it was necessary to select a sample of schools. All schools identified as having four or more first follow-up sample members enrolled were included in the school-level sample with certainty (probability = 1.0), and random samples were selected for retention from schools identified as having three first follow-up members (probability = 0.75), two first follow-up members (probability = 0.65), and one first follow-up member (probability = 0.31845). (Note that by the time of the data collection, only 1,380 of the 1,500 schools contained at least one NELS sample member.) Transcript and contextual data were requested for all students in the 1,380 selected schools.

In addition, transcripts were collected for all dropouts, early graduates, and 12th-grade sample members ineligible for the base-year, first follow-up, and second follow-up surveys owing to a language, physical, or mental barrier (triple ineligibles), through the sample "freshening" process, and the followback process of excluded students. This added 470 schools to the sample.

Of the 1,840 schools in the 1992 sample (including both contextual[1] and noncontextual schools), 1,540 participated in the 1992 study. Transcripts were requested for 19,320 students, and 17,290 transcripts were received.

The 1982 High School Transcript Study. The first transcript study was a component of the HS&B first follow-up. The 1982 study included students from 1,900 secondary schools—1,000 HS&B sampled schools and 900 schools to which students selected for the transcript survey had transferred (and for which no data collection activities other than transcript collection were carried out). Of these 1,900 schools, 1,720 provided transcripts. The total student sample size was 18,430 students. From the 1980 sophomores selected for the HS&B first follow-up, 12,310 cases were retained in the study sample with certainty—12,030 cases in the probability sample plus 280 nonsampled co-twins. In addition, a systematic sample of 6,120 cases was subsampled from the 17,700 remaining first follow-up selections, with a uniform probability of approximately .35. Transcripts were collected for 15,940 of the 18,430 students.

Data Collection and Processing

Data collection. The data collection and processing procedures are similar across the three transcript studies conducted as part of the Longitudinal Studies Program. The data collection procedures of the 2004 high school transcript study are discussed to illustrate the data collection process.

The ELS:2002 transcripts were collected from sample members in late 2004 and early 2005, about 6 months to 1 year after most students had graduated from high school. Collecting the transcripts in the 2004–05 school year allowed for more complete high school records. Transcripts were collected from the school that the students were originally sampled from in the base year (which was the only school for most sample members) and from their last school of attendance, if it was learned during the first follow-up student data collection that they had transferred. By requesting transcripts and related information for transfer students from a second school, this ELS:2002 transcript study offers the unique advantage of having extensive information on multiple school attendance and, therefore, increased accuracy of enrollmen histories. Incomplete records were obtained for sample members who had dropped out of school, had fallen behind the modal progression sequence, or were enrolled in a

[1] Schools selected for the contextual components of the second follow-up-the school administrator and teacher surveys-are referred to as contextual schools.

special education program requiring or allowing more than 12 years of schooling. For freshened students, transcripts were only collected from their senior year school. Transcripts were collected for regular graduates, dropouts, early graduates, and students who were homeschooled after their sophomore year.

From December 2004 through June 2005, survey materials were sent to over 2,000 schools. This group included schools that participated either in the base-year or first follow-up survey and transfer schools that were first contacted regarding ELS:2002 during transcript data collection. Transcripts were not requested from 10 base-year schools because they had refused to participate in the first follow-up survey. Additionally, transcripts were not requested from one base-year school that had no eligible students. Schools were paid $5 for each transcript. Transcripts were requested for over 16,000 sample members. Included were sample members who were ineligible to participate in the base year or first follow-up because of a physical disability, a mental disability, or a language barrier. Ninety-five schools required explicit consent from sample members or their parents/guardians before releasing transcript information. Of the sample members who attended these schools, about a quarter provided signed release forms. Two weeks after the survey materials were sent to the school, a follow-up postcard was sent as a reminder to complete the data collection forms and to send the requested materials to the Research Triangle Institute (RTI). If, after an additional week, RTI had not received the materials from the school, assigned institutional contactors (ICs) began telephone prompting to request that the materials be sent as soon as possible. Nonresponding schools contacted during the telephone prompting frequently requested remailing of the data collection materials. During telephone contacts, the ICs also identified any additional requirements the school had for releasing transcripts. Telephone follow-up with schools continued through June 2005. Additional measures were implemented to ensure an adequate response rate. In June 2005, data collection materials were sent to schools that had not yet provided all of the requested transcripts. In addition, in-person visits to nonresponding schools were conducted during April through June 2005 to collect the requested materials or to assist the school transcript preparer in assembling the information. For efficiency, the schools were selected for in-person visits by their proximity to other schools. In-person visits were made only to schools that had not sent transcript materials for any requested sample members.

Data processing. Each of the courses entered on the transcripts were coded using the Classification of Secondary School Courses (CSSC). The descriptions of the 2004 high school transcript data processing procedures illustrate the data processing done in the three transcript studies conducted as part of the Longitudinal Studies Program.

For the 2004 data processing, incoming data collection forms, transcripts, and course catalogs were logged into the survey control system by staff from RTI. Course catalog and transcript data were then entered using a web-based CADE system. Course catalogs from ELS:2002 base-year schools were keyed and coded for the preparation of course offerings data. For ELS:2002 base-year schools that provided them, courses listed in course catalogs were keyed and assigned the appropriate CSSC code before transcript keying and coding. For each catalog course entered, keyer-coders selected an appropriate course code from the CSSC look-up table in the data entry system. All transcripts received from a school were assigned to a single person for keying and coding. Course catalogs from non-base-year schools were not keyed. Data entry of each catalog and transcript was reviewed for accuracy by a supervisor or by a group of keyer-coders trained to perform these reviews. Procedures for editing, coding, error resolution, and documentation were modeled after the NELS:88 second follow-up transcript component (Ingels et al. 1995). Data entry systems included checks for valid variable ranges and codes, including legitimate missing codes, and CSSC code checks. Sequences of machine edits and visual data inspections were performed. Tasks included supplying missing data, detecting and correcting illegal codes, and investigating and resolving inconsistencies or anomalies in the data. Variable frequencies and cross-tabulations were reviewed to verify the correctness of machine editing.

Estimation Methods

Weighting. The weighting procedures used in the 2004 high school transcript study are presented as an example of the weighting procedures used in transcript studies conducted as part of the Longitudinal Studies Program.

In the 2004 high school transcript study, weight was assigned as follows. First, the first follow-up design weight was used as the starting weight. Next, Generalized Exponential Models (GEM) were used to compute weight adjustments. Weight adjustments included (1) a nonresponse adjustment to reduce potential bias owing to transcript nonresponse; and (2) a poststratification adjustment to ensure that sums for weights for certain domains had the same totals as those in the first follow-up. The nonresponse adjustment was performed in two stages: (1) at the

school refusal stage (e.g., the school refused to provide any transcript); and (2) at the within-school student-level nonresponse stage (see below for more details). Poststratification was performed to keep key estimates consistent with those in the first follow-up. Extreme weights were adjusted, truncated, and smoothed by GEM as part of the nonresponse and poststratification adjustments rather than as a separate step.

Imputation. Imputation was done for missing data in the High School Transcript Studies conducted for NELS and HS&B as part of the Longitudinal Studies Program.

Imputation was done for missing sex data in the 1992 NELS transcript study, using the student's first name to determine sex. In the 1982 HS&B transcript study, values were imputed for missing sex and race/ethnicity.

Data Quality and Comparability

Sampling Error

For the 1982, 1992, and 2004 high school transcript studies, variance estimation required the Taylor series linearization procedure, which took into account the complex sample design of these surveys, including stratification and clustering. This procedure takes the first-order Taylor series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. For stratified multistage surveys, the Taylor series procedure requires analysis strata and analysis PSUs (in ELS:2002, schools are the PSUs). Therefore, analysis strata and analysis PSUs were created in the base year and used again in the first follow-up.

Transcript studies conducted as part of the Longitudinal Studies Program may also use the Balanced Repeated Replication (BRR) variance estimation procedure or both Taylor Series linearization and BRR for variance estimation. For example, in NELS:88 and ELS:2002, variance estimation can be done in two ways: first, with Taylor Series linearization using software such as SUDAAN, AM, or STATA when using the Electronic Codebook (ECB) data; or, when using BRR, using the table generator (DAS—Data Analysis System) version of the dataset. Thus, the same estimate can have two different standard errors even within the same study, depending on whether its basis is a Taylor Series linearization or BRR. HS&B used both BRR and the Taylor Series and compared the results. These two methods result in very small differences that should not markedly change conclusions about the standard error of an estimate.

Coverage error. Potential sources of undercoverage in the high school transcript studies include (1) incomplete sampling frame data, as no national listing of schools is, or remains for very long, 100 percent complete and accurate; (2) omissions and errors in school rosters; and (3) deliberate exclusion of certain categories of students—such as students with physical or mental disabilities or non-English speakers, who might find it difficult or impossible to complete demanding cognitive tests and questionnaires. The first two sources are thought to have only a very small impact on high school transcript estimates. The most serious potential source is the undercoverage bias due to the exclusion of certain categories of students.

HS&B and NELS transcript studies are believed to exclude students with physical, mental, or linguistic barriers to assessment or survey participation. NELS transcript study collected data on the characteristics of excluded students, so that undercoverage bias can be quantified, and that the 1992 NELS study had negligible undercoverage of about 3 percent for the senior cohort. Although quantifiable exclusion data are not available for HS&B, given the similarity of eligibility rules in all two studies, it is reasonable to presume that HS&B exclusion rates were between 3 and 6 percent.

Nonresponse error.

Unit nonresponse. There is unit nonresponse at both the school and student levels in high school transcript studies. Response rates for all eligible high school transcript schools and students are presented in table 23.

Table 23. Unweighted response rates for all eligible High School Transcript Study schools and students in each study: 1982, 1992 an 2004

Year	School response rate	Student coverage rate
2004	79	91
1992	84	89
1982	91	88

SOURCE: Bozick, R., Lyttle, T., Siegel, P.H., Ingels, S.J., Rogers, J.E., Lauff, E., and Planty, M. (2006). *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation* (NCES 2006-338). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Transcripts were collected from 79 percent (unweighted) of the schools in the ELS:2004 study, 84 percent (unweighted) of the schools in the NELS:92 study, and 91 percent (unweighted) of the schools in the 1982 HS&B study.

At the student level, transcripts were obtained for 91 percent (unweighted) of eligible students in the ELS: 2004 study, 89 percent in the NELS:92 study, and 88 percent (unweighted) in the 1982 HS&B study.

Item nonresponse. Rates for item nonresponse have ranged from nonexistent to extremely high, depending on the type of item, across all of the high school transcript studies. As would be expected in transcript studies, course-level items have little if any nonresponse. Specific items include school year, term, and grade in which a course was taken; school-assigned course credits; and standardized course grade. However, nonresponse rates for items such as class size, cumulative GPA, class rank, days absent in each of the 4 high school years, and standardized test scores (e.g., PSAT, SAT, ACT) are very high.

In the 1992 NELS transcript study, the nonresponse rates for these items ranged from 0 percent for school year to less than 2 percent for the school term in which a course was taken. Incompleteness of actual course data, while considered to be limited, is another source of potential bias in a transcript study. Course data may be incomplete for students who transferred from one school to another. Also, it is difficult to assess the completeness of transcript data for dropouts in the 1982 HS&B and 1992 NELS transcript studies because of inconsistencies between enrollment reports of the sample member and the school.

Transcripts often provide other pieces of information that are useful in the analysis of coursetaking patterns: days absent in each school year, class rank, class size, month and year student left school, reason student left school (e.g., dropped out, graduated, transferred), cumulative GPA, participation in specialized courses or programs, and various standardized test scores (e.g., PSAT, SAT, ACT). While nonresponse rates for participation in specialized courses or programs (2 percent) and month/ year/reason student left school (less than 4 percent) are quite low in the 1992 NELS transcript study, nonresponse rates for the other items are very high: 18 percent for class size; 22 percent for cumulative GPA; 23 percent for class rank; 42–44 percent for days absent in each of the 4 high school years; and 67–73 percent for standardized test scores. (Note that although students were asked in a student questionnaire whether and when they planned to take specific tests, some students may not have actually taken the tests; this would, in part, explain the high nonresponse rates for test scores.

This wide range of item nonresponse rates is comparable to the range of nonresponse rates in the 1982 HS&B transcript study. For example, in the 1982 HS&B transcript study, the nonresponse rate was 32 percent for class rank and class size, 41–47 percent for days absent per school year, and 75 percent and above for standardized test scores.

Two key analytic variables are sex and race/ethnicity. Item nonresponse rates for sex have been extremely low: 0 percent in both the 1982 HS&B transcript study and the 1992 NELS transcript study. For race/ethnicity, nonresponse has ranged from 0 percent in the 1982 HS&B transcript study to 0.7 percent in the 1992 NELS transcript study.

Measurement error. Possible sources of measurement error in high school transcript studies are differences between schools and teachers in grading practices (e.g., grade inflation), differences in how data are recorded (although efforts are made to standardize grades and course credits for the high school transcript studies), and errors in keying or processing the transcript data (although the system has many built-in quality checks). The amount of measurement error in any survey or study is difficult to determine, and it is unknown for the high school transcript studies. However, because the transcripts are official school records of students' progress, it is reasonable to presume that there is less measurement error than in other types of data collections, particularly those that are self-reported.

Data Comparability

The high school transcript studies conducted by NCES have both similarities and dissimilarities of design and methodology that raise questions of comparability and may sometimes require analytical adjustments to ensure that comparability is maximized. This section presents four such issues: the comparability of target populations, sample inclusion and exclusion, methodology across studies, and content across studies. For details, please refer to the *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation* (Bozick et al. 2006).

*Comparability of target populations***.** The first comparability issue concerns the comparability of the target population. Comparable analysis samples can be achieved across the high school transcript studies by limiting analysis samples to high school graduates who received regular/standard or honors diplomas and

imposing additional restrictions such as earned credit minimums.

HS&B drew a national probability sample of high schools, as well as the sophomores and seniors within those schools, as of the 1980 spring term. By 1982, the school sample was no longer nationally representative (in the strictest sense) because it did not take into account school openings and closings in the 2-year period.

Similarly, while the HS&B senior cohort sample in 1980 generalized to the nation's high school seniors, the sophomore cohort in 1982 cannot be said to strictly represent the high school class of 1982. The HS&B sample was never freshened to add 1982 seniors who had no chance of selection 2 years before. This means that there is a bias in the HS&B 1982 (sophomores 2 years later) sample when it is used to generalize either to 12th-graders or to high school graduates. Seniors who were outside the United States 2 years before or seniors who were not sophomores 2 years before (e.g., seniors who repeated a year or who had a significantly accelerated trajectory) had no chance of selection into the sophomore cohort sample and are not represented within it.

The next two NCES high school cohort longitudinal studies, NELS:88 and ELS:2002, instituted a sample freshening procedure so that they include a nationally representative sample of high school seniors.

*Sample inclusion and exclusion***.** A second issue concerns student sample inclusion and exclusion, especially with respect to students with disabilities and English language learners.

In HS&B, sample members were classified as ineligible if deemed by their schools unable to complete the HS&B assessment battery owing to disability or lack of proficiency in English. Unfortunately, excluded students and specific reasons for exclusion were not well documented. However, it seems clear that the ineligible students represent the more severely disabled and the least proficient non-English speakers.

While some students were excluded from NELS:88, they were well documented, and over time their eligibility status was revisited. In ELS:2002, no students were excluded, though for those who could not complete survey forms, only contextual data and transcripts were collected. Also, in ELS:2002, some students received testing accommodations (e.g., extra time to complete the test); these cases are specially flagged.

Limiting 12th-grade high school graduate samples to recipients of regular or honors diplomas and eliminating cases that lack English course credits or that reflect a special education diploma or certificate of attendance largely eliminates the problem of differences in the excluded student population across studies. However, there is the remaining issue of how to identify and study the transcripts of individuals who had mild disabilities and how to compare the results over time. These issues arise because the longitudinal studies sought disability information from multiple respondent populations at multiple points in time. In NELS:88, for example, parents, teachers, students, and school administrators were all used as sources of information related to disability status.

Although some disability information is collected from sophomores' teachers, the primary source of identification for sophomore cohort members with disabilities in ELS:2002 is the Individualized Educational Program (IEP) flag, based on information taken from the sampling records provided by the base-year school, which identifies students in the school with IEPs.

Methodology across studies. In addition to differences in target populations and inclusion criteria, there are other differences among NCES high school transcript studies in terms of methodology. First, there is some variation in the statistical procedures used across studies. Overall, this variation will be the source of small differences that should not disrupt trend analyses. For example, different methods were used for nonresponse adjustment of weights. In HS&B, weighting cells were constructed based upon the known characteristics of the sample units. ELS:2002 used propensity modeling rather than a weighting cell approach. In NELS:88, a mix of the two approaches is encountered (propensity at the school level, weighting cells at the student level). However, results of nonresponse adjustment tend to be highly correlated regardless of method. Therefore, these differences should not lead to greatly different estimates.

*Content across studies***.** As curriculum changes, new courses emerge while others fall by the wayside. Therefore, with every transcript study, there is a need to add courses to the CSSC. Additionally, SST has been revised twice to accommodate changes in the curriculum. From a classification standpoint, adding new subject areas (such as information processing and computer studies) and expanded course offerings (including more AP courses) presents less of a quandary than certain efforts to achieve curriculum integration through interdisciplinary courses—confining such offerings (e.g., history of mathematics,

philosophy of science, psychological anthropology) in one subject category does injustice to certain aspects of the course content, while counting such courses in multiple areas may magnify and distort their impact.

As a result of these changes, many transcript composite variables have also changed over time. For example, with initiatives to seamlessly integrate academic into vocational education, conceptualizations of track or program type have changed. Such differences may reduce ease and simplicity in trend analysis, but are unavoidable features of the need to confront a complex and changing reality. Also, HS&B did not use as refined a system of course classification as did later studies, which, for example, distinguished courses based on whether they were remedial, regular, or advanced. On the other hand, some new measures developed out of NELS:88, such as the “pipeline” variables, which measure course content level, can be “read into” the other studies, such as the HS&B transcript studies.

The major limitation of these changes is that there are few coursetaking variables that are directly comparable across studies. For example, only a handful of courses qualified as computer science in the HS&B study. As the number of computer science courses has expanded, any variable based on computer science is not truly comparable across studies because it does not capture the range of courses that have emerged over time. Along with the two revisions of the SST, these changes make direct comparisons among coursetaking variables in the different files difficult. To facilitate some comparisons, ELS:2002 provides six summary measures that have directly comparable variables in NELS:88 and that can be constructed in HS&B by using existing elements. These variables are based on the same CSSC codes.

Analysts interested in comparing coursetaking patterns need to examine the CSSC codes available in each study. The CSSC codes are the same across studies, thus facilitating direct comparisons. As noted earlier, the list has evolved and certain subject areas have changed accordingly. Users may want to construct measures in a variety of ways to ensure that their findings are robust with respect to different variable specifications. In addition, analysts should consider changes in subject areas over time when conducting time trend analyses and interpreting findings.

There are many other variables that are typically linkable to transcripts; however, their status for this purpose may sometimes be problematic. For example, in HS&B and NELS:88, race was self-reported and students were asked to mark only one race. In light of the 2000 decennial census and revised race-reporting guidelines issued by the Office of Management and Budget, a new race category was added at the time of ELS:2002. More importantly, ELS:2002 respondents were allowed to mark all applicable races, thus generating a further category—multiracial. Knowing if a respondent who self-identified as Black on the HS&B questionnaire would have self-identified only as Black on the ELS:2002 questionnaire is impossible. To this extent, coursetaking trends for Blacks will be more uncertain than if a consistent definition had been maintained.

Test scores are another set of variables typically linked with transcript data that are different across studies. The relationship between coursetaking and tested achievement is of interest to researchers, and exploring the relationship between curriculum and assessment results is an interesting area for time series analysis. The NCES transcript studies provide only limited scope for such explorations.

Between NLS:72 in 1972 and ELS:2002 in 2004, the only subject consistently tested was mathematics. A further complication with comparative use of assessment data is changes in the measurement scale. Selectively, where content similarities permit, this limitation has been overcome by test linkage, usually IRT-based or equipercentile equating. One could, for example, examine the relationship between coursetaking and gain in the first 2 years of high school, using the equated 1980, 1990, and 2002 mathematics scores, or one could examine the relationship between coursetaking and gain for the periods 1990–92 and 2002–04, since ELS:2002 has been put on the NELS:88 scale. One final option for use of assessment data is to examine change within an effect size metric.

Contact Information

For content information about the High School Transcript Studies conducted as part of the Longitudinal Studies Program, contact

Jeffrey Owings
Phone: (202) 502-7423
E-mail: jeffrey.owings@ed.gov

Mailing Address:
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

Methodology and Evaluation Reports

Alt, M.N., and Bradby, D. (1999). *Procedures Guide for Transcript Studies* (NCES Working Paper 1999-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Bozick, R., Lyttle, T., Siegel, P.H., Ingels, S.J., Rogers, J.E., Lauff, E., and Planty, M. (2006). *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation* (NCES 2006-338). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Bradby, D., and Hoachlander, G. (1999). *1998 Revision of the Secondary School Taxonomy* (NCES Working Paper 1999-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Ingels, S.J., Dowd, K.L., Taylor, J.R., Bartot, V.H., Frankel, M.R., Pulliam, P.A., and Quinn, P. (1995). *National Education Longitudinal Study of 1988 Second Follow-Up: Transcript Component Data File User's Manual* (NCES 95-377). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Planty, M., Bozick, R., and Ingels, S.J. (2006). *Academic Pathways, Preparation, and Performance: A Descriptive Overview of the Transcripts from the High School Graduating Class of 2003-04* (NCES 2007-316). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chapter 30: Quick Response Information System

NCES has established two survey systems to collect time-sensitive, issue-oriented data quickly and with minimal response burden. The Fast Response Survey System focuses on collecting data at the elementary and secondary school levels. The Postsecondary Education Quick Information System collects data at the postsecondary level. These systems, subsumed under the general title, Quick Response Information System, are used to meet the data needs of U.S. Department of Education analysts, planners, and decisionmakers when information cannot be obtained quickly through traditional National Center for Education Statistics (NCES) surveys.

TWO QUICK RESPONSE INFORMATION SYSTEMS FAST RESPONSE SYSTEMS:

- Fast Response Survey System (FRSS) – 100 surveys since 1975
- Postsecondary Education Quick Information System (PEQIS) – 16 surveys since 1991

Fast Response Survey System

Overview

The Fast Response Survey System (FRSS) was established in 1975 to collect issue-oriented data quickly and with minimum response burden. The FRSS, whose surveys collect and report data on key education issues, was designed to meet the data needs of U.S. Department of Education analysts, planners, and decisionmakers when information could not be collected quickly through NCES's large recurring surveys. Findings from FRSS surveys have been included in congressional reports, testimony to congressional subcommittees, NCES reports, and other Department of Education reports. The findings are also often used by state and local education officials. From 1975 to 1990, the FRSS collected data at all education levels. Since the Postsecondary Education Quick Information System (PEQIS) was established in 1991, FRSS surveys have been limited to elementary and secondary school issues. To date, some 100 surveys have been conducted under the FRSS. Topics have ranged from racial and ethnic classifications at the state and school levels to the availability and use of resources, such as advanced telecommunications and libraries. Additionally, data have been collected on education reform, violence and discipline problems, parental involvement, curriculum placement and arts education, nutrition education, teacher training and professional development, vocational education, children's readiness for school, and the perspectives of school district superintendents, principals, and teachers on safe, disciplined, and drug-free schools. Some surveys, like surveys on Internet access and on teacher preparation and qualifications, have been conducted more than once in the past.

Data from FRSS surveys are representative at the national level, drawing from a universe that is appropriate for each study. Since 1991, the FRSS has generally collected data from public and private elementary and secondary schools, elementary and secondary school teachers and principals, public and school libraries, and, less frequently, state education agencies and local education agencies. Prior to 1991, FRSS also collected data from postsecondary institutions.

Sample Design

Data collected through FRSS surveys are representative at the national level, drawing from a universe that is appropriate for each study.

The FRSS collects data from state education agencies and national samples from other education organizations and participants, including local education agencies, public and private elementary and secondary schools, , and elementary and secondary school teachers and principals. To ensure minimal burden on respondents, the surveys are generally limited to three pages of questions, with a response burden of about 30 minutes per respondent. Sample sizes are relatively small (usually about 1,200 to 1,500 respondents per survey, but occasionally larger) so that data collection can be completed quickly.

The sampling frame for FRSS surveys is typically the NCES Common Core of Data (CCD) public school (or agency) universe. (See chapter 2.) The following variables are usually used for stratification or sorting within primary strata: instructional level (elementary school, middle school, and high school [secondary/combined]); categories of enrollment size; locale (city, urban fringe, town, rural); geographical region (Northeast, Southeast, Central, West); and categories of poverty status (based on eligibility for free or reduced-price lunch). The allocation of the samples to the primary strata is intended to ensure that the sample sizes are large enough to permit analyses of the questionnaire for major subgroups.

Within primary strata, the sample sizes are frequently allocated to the substrata in rough proportion to the aggregate square root of the size of enrollment of schools in the substratum. The use of the square root of enrollment to determine the sample allocation is considered reasonably efficient for estimating school-level characteristics and quantitative measures correlated with enrollment.

For example, the sample of elementary and secondary/combined schools for *Educational Technology in U.S. Public Schools: Fall 2008* was selected from the 2005–06 CCD Public School Universe data file, the most up-to-date file available at the time the sample was drawn. The sampling frame included over 85,000 regular schools. Excluded from this sampling frame were schools with a high grade of prekindergarten or kindergarten and ungraded schools, along with special education, vocational, and alternative/other schools; schools outside the 50 states and the District of Columbia; and schools with zero or missing enrollment.

The public school sampling frame was stratified by level (elementary or secondary/combined), categories of enrollment size, and categories for percent of students eligible for free/reduced-price lunch. Schools in the frame were then sorted by locale and region to induce additional implicit stratification. A sample of 2,010 schools were selected for the sample, but 56 were found to be ineligible for the survey because they were closed, merged, or did not meet the eligibility requirements for inclusion (e.g., they were special education, vocational, or alternative schools). This left a total of 1,950 eligible schools in the sample.

FRSS survey samples are sometimes constructed from the NCES Private School Universe Survey (PSS). (See chapter 3.) The sample usually consists of regular private elementary, secondary, and combined schools, with a private school being defined as a school not in the public system that provides instruction for any of grades 1–12 (or comparable ungraded levels) where the instruction is not provided in a private home. The following variables may be used for stratification or sorting within primary strata: instructional level (elementary, secondary, and combined), affiliation (Catholic, other religious, and nonsectarian), school size, geographic region, locale, and percentage of Black, Hispanic, and other race/ethnicity students. Schools are generally selected from each primary stratum with probabilities proportional to the weight reflecting the school's probability of inclusion in the area sample.

Other sources may serve as sampling frames, depending on the needs of the survey. For example, for *Participation of Migrant Students in Title I Migrant Education Program (MEP) Summer-Term Projects*, the districts and other entities serving migrant students were selected from the U.S. Department of Education's 1995–96 Migrant Education Program Universe data file.

Some FRSS surveys use a two-stage sampling process. For example, the *Teachers' Use of Educational Technology in U.S. Public Schools: 2009* and the *Educational Technology in Public School Districts: Fall 2008* which were administered concurrently with the *Educational Technology in U.S. Public Schools: Fall 2008* had a two-stage sampling process. The schools were selected during the first stage. The second stage of sampling for the Teacher Survey involved obtaining lists of teachers from the selected schools. The second stage of sampling for the Public School District Survey identified the districts that contained at least one of the sampled schools using the 2005-06 CCD Local Education Agency file.

Before PEQIS was established, the FRSS was sometimes used to examine postsecondary issues. For example, the *College-Level Remedial Education in the Fall of 1989* targeted institutions of higher education that served freshmen and were accredited at the college

level by an association or agency recognized by the U.S. Secretary of Education. The sampling frame was the universe file of the Higher Education General Information System (HEGIS) Fall Enrollment and Compliance Report of Institutions of Higher Education of 1983–84. (Note that HEGIS has since been replaced by the Integrated Postsecondary Education Data System—IPEDS—see chapter 12.) The universe of colleges and universities was stratified by type of control, type of institution, and enrollment size. Within strata, schools were selected at uniform rates, but the sampling rates varied considerably from stratum to stratum.

Data Collection and Processing

Most FRSS surveys are self-administered questionnaires where respondents are offered the option of completing the survey by mail or via the Web, with telephone follow-up for survey nonresponse and data clarification. On rare occasion a few have been telephone surveys, including one that used random digit dialing techniques. FRSS questionnaires are pretested, and efforts are made to check for consistency in the interpretation of questions and to eliminate ambiguous items before fielding the survey. For example, for the *Educational Technology in Public School Districts: Fall 2008* survey, questionnaires and cover letters were mailed to the superintendent of each sampled school district in early August 2008. The letter introduced the study and requested that the questionnaire be completed by the person most knowledgeable about educational technology in the district. Respondents were offered the option of completing the survey by mail or via the Web. Telephone follow-up for survey nonresponse and data clarification was initiated in late August 2008 and completed in January 2009.

Data are keyed with 100 percent verification. To check the data for accuracy and consistency, questionnaire responses undergo both manual and machine editing. Cases with missing or inconsistent items are recontacted by telephone.

Westat has served as the contractor for all surveys.

Estimation

Weighting. The response data are weighted to produce national estimates. The weights are designed to adjust for the variable probabilities of selection and differential nonresponse. Out-of-scope units are deleted from the initial sample before weighting and analysis. In the case of two-stage sampling—for example, in the *Teachers' Use of Educational Technology in U.S. Public Schools: 2009* —the weights used to produce national estimates were designed to reflect the variable probabilities of selection of the sampled schools and teachers and were adjusted for differential unit (teacher sampling list and questionnaire) nonresponse.

Imputation. Because item nonresponse rates in FRSS surveys are typically very low, the use of imputation is limited. The missing data are imputed using a "hot-deck" approach to obtain a "donor" from which the imputed values are derived. Once a donor is found, it is used to derive the imputed values for the missing data. For categorical items, the imputed value is simply the corresponding value from the donor. For numerical items, an appropriate ratio (e.g., the proportion of instructional rooms with wireless internet connections) is calculated for the donor, and this ratio is applied to available data (e.g., reported number of instructional rooms) for the recipient to obtain the corresponding imputed value. All missing items for a recipient are imputed from the same donor.

For example, in the *Educational Technology in U.S. Public Schools: Fall 2008* survey, all questionnaire items with response rates of less than 100 percent were imputed using the hot-deck imputation method. Under the "hot-deck" approach, a "donor" school that matched selected characteristics of the school with missing data (the recipient school) was identified. This survey used instructional level, categories of enrollment size, region, categories for percent combined enrollment of Black, Hispanic, Asian/Pacific Islander, or American Indian/Alaska Native students, categories for percent of students in the school eligible for free or reduced-price lunch, district size, and district poverty level as the matching characteristics. In addition, relevant questionnaire items were used to form appropriate imputation groupings. Once a donor was found, it was used to obtain the imputed values for the school with missing data. For categorical items, the imputed value was simply the corresponding value from the donor school. For the numerical items, an appropriate ratio was calculated for the donor school, and this ratio was applied to available data for the recipient school to obtain the corresponding imputed value.

Sampling Error

FRSS estimates are based on the selected samples and, consequently, are subject to sampling variability. The standard error is a measure of the variability of estimates due to sampling. Jackknife replication is the method used to compute estimates of the standard errors.

Nonsampling Error

Coverage Error. FRSS surveys are subject to any coverage error present in the major NCES data files

that serve as their sampling frames. Many FRSS surveys use CCD surveys as the sampling frame.

There is a potential for undercoverage bias associated with the absence of schools built between the time when the sampling frame is constructed and the time of the FRSS survey administration. Since teacher coverage depends on teacher lists sent by the schools, teacher coverage is assumed to be good. (See chapter 2 for a description of the CCD; see relevant chapters for other NCES surveys that serve as sampling frames for FRSS surveys.)

Nonresponse Error. Unit response for most FRSS surveys is 90 percent or higher. (See table 23.) Item nonresponse for most items is less than 1 percent. The weights are adjusted for unit nonresponse.

Measurement Error. Errors may result from such problems as misrecording of responses; incorrect editing, coding, and data entry; different interpretations of definitions and the meaning of questions; memory effects; the timing of the survey; and the respondent's inability to report certain data due to deficiencies in a recordkeeping system. Several specific examples of possible measurement error come from *the Public School Survey on Education Reform* and the *Public School Teacher Survey on Education Reform*, conducted in 1996. Survey results should be interpreted carefully for the following reasons: (1) survey questions were designed to be inclusive of a wide variety of reform activities since all principals and teachers do not share the same concept of reform; (2) respondents may have overreported activities in which they believe they should have been engaged; and (3) the questionnaire was too brief to collect information that could assist in judging the accuracy of the respondents' reports.

Data Quality and Comparability

Some FRSS surveys, such as surveys on internet access and on teacher preparation and qualifications, are repeated so that results can be compared over time. For example, the FRSS conducted the *Survey on Advanced Telecommunications in U.S. Public Schools* in 1994, 1995, 1996, and 1997. More recently, *Internet Access in U.S. Public Schools and Classrooms* was administered in 1998, 1999, 2000, 2001, 2002, 2003, and 2005. In addition, the *Survey on Advanced Telecommunications in U.S. Private Schools* was administered in 1995 and 1998–99. Results from the 1997 *Principal/School Disciplinarian Survey on School Violence* can be compared with those from the 1991 *Principal Survey on Safe, Disciplined, and Drug-Free Schools*, although there are some sampling differences that should be taken into account. (The 1997 survey was restricted to regular elementary and secondary schools, whereas the 1991 survey also included 13 vocational education and alternative schools in the sample.) Another example is provided by *Technology-Based Distance Education Courses for Public Elementary and Secondary School Students*, which was administered in 2002–03 and 2004–05. Two types of comparisons are possible with these FRSS data. The first type involves comparisons of the cross-sectional estimates for the two or more time periods. Cross-sectional comparisons reflect the net change in a given characteristic across years, including any changes in the underlying population. However, the enrollment estimates for 2002–03 and 2004-05 are different due to extensive data quality control procedures in place during data collection for the 2004–05 survey. The second type of comparison provides longitudinal analysis of change between 2002–03 and 2004–05. The longitudinal analysis is based on data from both administrations of the distance education survey, with the districts that existed both in 2002–03 and 2004–05 included in the analysis.

Occasionally, an FRSS survey is fielded to provide data that can be compared with data from another NCES survey. For example, the 1996 *Survey on Family and School Partnerships in Public Schools, K–8* was designed to provide data that could be compared with parent data from the 1996 National Household Education Survey as well as with data from the Prospects Study, a congressionally mandated study of educational growth and opportunity from 1991 to 1994. Another example is the 2001 *Survey on High School Guidance Counseling*, which was designed to provide data that could be compared to data from the 1984 Administrator and Teacher Survey supplement to the High School and Beyond Longitudinal Study.

Contact Information

For content information about the FRSS project, contact:

Peter C. Tice
Phone: (202) 502-7497
E-mail: peter.tice@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

Table 24. Weighted unit response rates for recent FRSS surveys: Selected years, 1999–2010

Survey	List participation rate	Weighted 1st level response rate	Overall weighted response rate
Teachers' Use of Educational Technology in U.S. Public Schools, 2009	81	79	65
Educational Technology in Public School Districts, Fall 2008	†	90	90
Educational Technology in U.S. Public Schools, Fall 2008	†	79	79
After-School Programs in Public Elementary Schools, 2008	†	91	91
Alternative Schools and Programs for Public School Students At Risk of Educational Failure, 2007-08	†	96	96
Distance Education Courses for Public School Elementary and Secondary School Students: 2004–05	†	96	96
Foods and Physical Activity in Public Elementary Schools: 2005	†	91	91
Public School Principals' Perceptions of Their School Facilities: Fall 2005	†	91	91
Internet Access in U.S. Public Schools and Classrooms: Fall 2005	†	86	86
Internet Access in U.S. Public Schools and Classrooms: Fall 2003	†	92	92
Internet Access in U.S. Public Schools and Classrooms: Fall 2002	†	90	90
Dual Credit and Exam-Based Courses: 2003	†	92	92
Distance Education Courses for Public School Elementary and Secondary School Students: 2002–03	†	96	96
Effects of Energy Needs and Expenditures on U.S. Public Schools: 2001	†	84	84
Survey on High School Guidance Counseling: 2001	†	94	94
District Survey of Alternative Schools and Programs: 2001	†	97	97

See notes at end of table.

Table 24. Weighted unit response rates for recent FRSS surveys: Selected years, 1999–2010–Continued

Survey	List participation rate	Weighted 1st level response rate	Overall weighted response rate
Survey of Classes that Serve Children Prior to Kindergarten in Public Schools: 2000–01	†	94	94
Survey on Programs for Adults in Public Library Outlets: 2000	†	97	97
Survey on Professional Development and Training in U.S. Public Schools: 1999–2000	88	85	75

† Not applicable.

SOURCE: Carver, P.R., and Lewis, L. (2010). *Alternative Schools and Programs for Students At Risk of Educational Failure, 2007-08* (NCES 2010-026). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Chaney, B., and Lewis, L. (2007). *Public School Principals' Report on Their School Facilities: Fall 2005* (NCES 2007-007). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Gray, L., and Lewis, L. (2009). *Educational Technology in Public School Districts, Fall 2008* (NCES 2010-003). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Gray, L., Thomas, N., and Lewis, L. (2010). *Educational Technology in U.S. Public Schools, Fall 2008* (NCES 2010-034). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Gray, L., Thomas, N., and Lewis, L. (2010). *Teachers' Use of Educational Technology in U.S. Public Schools, 2009* (NCES 2010-040). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Kleiner, A., and Lewis, L. (2003). *Internet Access in U.S. Public Schools and Classrooms: 1994–2002* (NCES 2004-011). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Kleiner, B., Porch, R., and Farris, E. (2002). *Public Alternative Schools and Programs for Students at Risk of Education Failure: 2000–01* (NCES 2002-004). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Lewis, L., and Farris, E. (2002). *Programs for Adults in Public Library Outlets* (NCES 2003-010). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Parsad, B., and Jones, J. (2005). *Internet Access in U.S. Public Schools and Classrooms: 1994–2003* (NCES 2005-015). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Parsad, B., and Lewis, L. (2006). *Calories In, Calories Out: Food and Exercise in Public Elementary Schools, 2005* (NCES 2006-057). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Parsad, B., and Lewis, L. (2009). *After-School Programs in Public Elementary Schools* (NCES 2009-043). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Parsad, B., Lewis, L., and Farris, E. (2001). *Teacher Preparation and Professional Development: 2000* (NCES 2001-088). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Setzer, J.C., and Lewis, L. (2005). *Distance Education Courses for Public Elementary and Secondary School Students: 2002–03* (NCES 2005-010). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Smith, T., Kleiner, A., Parsad, B., and Farris, E. (2003). *Prekindergarten in U.S. Public Schools: 2000–2001* (NCES 2003-019). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Smith, T., Porch, R., Farris, E., and Fowler, W. (2003). *Effects of Energy Needs and Expenditures on U.S. Public Schools* (NCES 2003-018). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Waits, T., Setzer, J.C., and Lewis, L. (2005). *Dual Credit and Exam-Based Courses in U.S. Public High Schools: 2002–03* (NCES 2005-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Wells, J., and Lewis, L. (2006). *Internet Access in U.S. Public Schools and Classrooms: 1994–2005* (NCES 2007-020). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Zandberg, I., and Lewis L. (2008). *Technology-Based Distance Education Courses for Public Elementary and Secondary School Students: 2002–03 and 2004–05* (NCES 2008-008). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Methodology and Evaluation Reports

Methodology is discussed in the technical notes to survey reports. Some of these reports are listed below.

Alexander, D., Heaviside, S., and Farris, E. (1999). *Status of Education Reform in Public Elementary and Secondary Schools: Teachers' Perspectives* (NCES 1999-045). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Carey, N., Lewis, L., and Farris, E. (1998). *Parent Involvement in Children's Education: Efforts by Public Elementary Schools* (NCES 98-032). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Carpenter, J. (1992). *Public School District Survey on Safe, Disciplined, and Drug-Free Schools* (NCES 92-008). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Carver, P.R., and Lewis, L. (2010). *Alternative Schools and Programs for Students At Risk of Educational Failure, 2007-08* (NCES 2010-026). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Chaney, B., and Lewis, L. (2007). *Public School Principals' Report on Their School Facilities: Fall 2005* (NCES 2007-007). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Gray, L., and Lewis, L. (2009). *Educational Technology in Public School Districts, Fall 2008* (NCES 2010-003). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Gray, L., Thomas, N., and Lewis, L. (2010). *Educational Technology in U.S. Public Schools, Fall 2008* (NCES 2010-034). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Gray, L., Thomas, N., and Lewis, L. (2010). *Teachers' Use of Educational Technology in U.S. Public Schools, 2009* (NCES 2010-040). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hamann, T. (2000). *Coverage Evaluation of the 1994-95 Common Core of Data: Public Elementary/Secondary School Universe Survey* (NCES 2000-12). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Jackson, B., and Frazier, R. (1996). *Improving the Coverage of Private Elementary-Secondary Schools* (NCES 96-26). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kleiner, A., and Lewis, L. (2003). *Internet Access in U.S. Public Schools and Classrooms: 1994–2002* (NCES 2004-011). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kleiner, B., Porch, R., and Farris, E. (2002). *Public Alternative Schools and Programs for Students at Risk of Education Failure: 2000–01* (NCES 2002–004). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Lewis, L., and Farris, E. (2002). *Programs for Adults in Public Library Outlets* (NCES 2003-010). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Owens, S. (1997). *Coverage Evaluation of the 1994-95 Common Core of Data: Public Elementary/ Secondary Education Agency Universe Survey* (NCES 97-505). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Parsad, B., Alexander, D., Farris, E., and Hudson, L. (2003). *High School Guidance Counseling* (NCES 2003-015). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Parsad, B., and Jones, J. (2005). *Internet Access in U.S. Public Schools and Classrooms: 1994–2003* (NCES 2005-015). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Parsad, B., and Lewis, L. (2006). *Calories In, Calories Out: Food and Exercise in Public Elementary Schools, 2005* (NCES 2006-057). National Center for Education Statistics, Institute of Education

Sciences, U.S. Department of Education. Washington, DC.

Parsad, B., and Lewis, L. (2009). *After-School Programs in Public Elementary Schools* (NCES 2009-043). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Parsad, B., Lewis, L., and Farris, E. (2001). *Teacher Preparation and Professional Development: 2000* (NCES 2001-088). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Setzer, J.C., and Lewis, L. (2005). *Distance Education Courses for Public Elementary and Secondary School Students: 2002–03* (NCES 2005-010). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Smith, T., Kleiner, A., Parsad, B., and Farris, E. (2003). *Prekindergarten in U.S. Public Schools: 2000–2001* (NCES 2003-019). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Smith, T., Porch, R., Farris, E., and Fowler, W. (2003). *Effects of Energy Needs and Expenditures on U.S. Public Schools* (NCES 2003-018). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Waits, T., Setzer, J.C., and Lewis, L. (2005). *Dual Credit and Exam-Based Courses in U.S. Public High Schools: 2002–03* (NCES 2005-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wells, J., and Lewis, L. (2006). *Internet Access in U.S. Public Schools and Classrooms: 1994–2005* (NCES 2007-020). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Zandberg, I., and Lewis L. (2008). *Technology-Based Distance Education Courses for Public Elementary and Secondary School Students: 2002–03 and 2004–05* (NCES 2008-008). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Postsecondary Education Quick Information System

Overview

The Postsecondary Education Quick Information System (PEQIS) was established in 1991 to quickly collect limited amounts of policy-relevant information from a nationally representative sample of postsecondary institutions. Policy analysts, program planners, and decisionmakers in postsecondary education frequently need data on emerging issues quickly. It is not always feasible for NCES to use its large, recurring surveys to provide such data quickly, due to the length of time required to implement large-scale data collection efforts. In addition to obtaining information on emerging issues quickly, PEQIS surveys are used to assess the feasibility of developing large-scale data collection efforts on a given topic or to supplement other NCES postsecondary surveys. Surveys are generally limited to three pages of questions, with a response burden of about 30 minutes per respondent. To date, 16 PEQIS surveys have been completed, covering such diverse issues as distance learning, precollegiate programs for disadvantaged students, remedial education, campus crime and security, finances, services for deaf and hard-of-hearing students, and the accommodation of disabled students.

Sample Design

PEQIS employs a standing sample (panel) of approximately 1,600 nationally representative postsecondary education institutions at the 2- and 4-year levels. The panel includes public and private colleges and universities that award associate's, bachelor's, master's, and doctoral degrees. PEQIS can also conduct surveys of state higher education agencies. Four panels have been recruited since PEQIS was established in 1991. The sampling frame for the first PEQIS panel, recruited in 1992, was the 1990–91 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) file. (See chapter 12.) The sampling frame for the second PEQIS panel, recruited in 1996, was the 1995–96 IPEDS IC file. The PEQIS panel was reselected in 1996 to reflect changes in the postsecondary education universe since the 1992 panel was recruited. A modified Keyfitz approach was used to maximize overlap between the 1992 and 1996 panels; this resulted in 80 percent of the institutions in the 1996 panel overlapping with the 1992 panel. The sampling frame for the third PEQIS panel, recruited in 2002, was the 2000 IPEDS IC file. A modified Keyfitz approach was used to maximize the overlap between the 1996 and 2002 samples; 81 percent of the institutions overlapped between these

two panels. The sampling frame for the 2006 PEQIS panel was the 2005 IPEDS IC file. The modified Keyfitz approach used to maximize the overlap between the 2002 and 2006 panels resulted in 79 percent of the institutions overlapping between the two panels.

Institutions eligible for the PEQIS frames for the 1992 and 1996 panels included 2-year and 4-year (including graduate-level) postsecondary institutions and less-than-2-year institutions of higher education. In 2002 and 2006, institutions eligible for the PEQIS frames were 2-year and 4-year (including graduate-level) Title IV eligible, degree-granting postsecondary institutions. In 1992, the sampling frame covered the 50 states, the District of Columbia, and Puerto Rico. In 1996 and subsequent years, institutions in Puerto Rico were excluded. The sampling frame included 5,320 institutions in 1992; 5,350 institutions in 1996; 4,180 institutions in 2002; and 4,270 institutions in 2006.

The sampling frames for all four PEQIS panels were stratified by instructional level (4-year and 2-year institutions for all four panels plus less-than-2-year institutions for the 1992 and 1996 panels); control (public, private nonprofit, private for-profit); highest level of offering (doctor’s/first professional, master’s, bachelor’s, less than bachelor’s); and total enrollment. Within each of the strata, institutions were sorted by region (Northeast, Southeast, Central, West), whether the institution had a relatively high percentage of Black, Hispanic, and other race/ethnicity students; and, in 1992 and 1996 only, whether the institution had research expenditures exceeding $1 million. The 1992 sample of 1,670 institutions was allocated to the strata in proportion to the aggregate square root of *full-time-equivalent enrollment.* The 1996 sample of 1,670 institutions was allocated to the strata in proportion to the aggregate square root of *total enrollment*, as was the 2002 sample of 1,610 institutions and the 2006 sample of 1,630 institutions. For all four panels, institutions within a stratum were sampled with equal probabilities of selection.

During recruitment for the 1992 panel, 50 institutions were found to be ineligible for PEQIS, primarily because they had closed or offered just correspondence courses. The final unweighted response rate at the end of PEQIS panel recruitment in spring 1992 was 98 percent (1,580 of the 1,620 eligible institutions). The weighted response rate for panel recruitment (weighted by the base weight) was 96 percent.

The modified Keyfitz approach used in 1996 resulted in 80 percent of the institutions in the 1996 panel overlapping with the 1992 panel. Panel recruitment was conducted with the 340 institutions that were not part of the overlap sample. Twenty institutions were found to be ineligible for PEQIS. The final unweighted response rate for the institutions that were not part of the overlap sample was 98 percent. The final participation rate across all 1,670 institutions selected for the 1996 panel was about 100 percent. The weighted panel participation rate (weighted by the base weight) was about 100 percent.

The modified Keyfitz approach used in 2002 resulted in 81 percent of the institutions in the 2002 panel overlapping with the 1996 panel. Panel recruitment was conducted with the 300 institutions that were not part of the overlap sample. During panel recruitment, 6 institutions were found to be ineligible for PEQIS. The final unweighted response rate at the end of PEQIS panel recruitment with the institutions that were not part of the overlap sample was 97 percent. There were 1,600 eligible institutions in the entire 2002 panel, because 4 institutions in the overlap sample were determined to be ineligible for various reasons. The final unweighted participation rate across the institutions selected for the 2002 panel was 99 percent (1,590 participating institutions out of 1,600 eligible institutions). The weighted panel participation rate was also 99 percent.

The modified Keyfitz approach used in 2006 resulted in 79 percent of the institutions in the 2006 panel overlapping with the 2002 panel. Panel recruitment was conducted with the 340 institutions selected for the 2006 panel that were not part of the 2002 panel. During panel recruitment, some institutions were found to be ineligible for PEQIS because they had closed. The final unweighted response rate at the end of PEQIS panel recruitment with the institutions that were not part of the overlap sample was 86 percent (290 of the 340 eligible institutions). There were 1,620 eligible institutions in the entire 2006 panel. The final unweighted participation rate across the institutions selected for the 2006 panel was 97 percent (1,570 participating institutions out of 1,620 eligible institutions). The weighted panel participation rate was 93 percent.

Data Collection and Processing

Typically, PEQIS surveys are self-administered questionnaires with respondents offered the option of completing the survey by mail or via the Web, with telephone follow-up for survey nonresponse and data clarification. Surveys are limited to three pages of questions, with a response burden of about 30 minutes per respondent. The questionnaires are pretested, and efforts are made to check for consistency in the interpretation of questions and to eliminate ambiguous

items before fielding the survey to all institutions in the sample.

The questionnaires are sent to institutional survey coordinators who identify the appropriate respondents for the particular survey and forward questionnaires to them. Nonrespondents who have not returned the survey within a set period of time are followed up by telephone. Data are keyed with 100 percent verification. To check the data for accuracy and consistency, questionnaire responses undergo both manual and machine editing. Cases with missing or inconsistent items are recontacted by telephone.

For the *Distance Education at Degree-Granting Postsecondary Institutions: 2006-07* survey, questionnaires were mailed to the PEQIS coordinators at the 1,630 institutions in fall 2007. The coordinators were told that the survey was designed to be completed by the person at the institution most knowledgeable about its distance education programs. In addition, data were collected from one 4-year private for-profit institution that was added to the sample for this survey only because it is the largest provider of online distance education courses in the nation, bringing the total sample size for this survey to 1,630 institutions. Respondents had the option of completing the survey online. Telephone follow-up of nonrespondents was initiated 3 weeks after mailout; data collection and clarification were completed in March 2008. Of the institutions that completed the survey, 72 percent completed it online, 20 percent completed it by mail, 5 percent completed it by fax, and 4 percent completed it by telephone.

Westat has served as the contractor for all surveys.

Weighting

The response data are weighted to produce national estimates. The weights are designed to adjust for the variable probabilities of selection and differential nonresponse. For recent PEQIS surveys, the weighted number of eligible institutions represents the estimated universe of approximately 4,240 Title IV-eligible degree-granting institutions in the 50 states and the District of Columbia.

Imputation

Item nonresponse rates for PEQIS surveys are typically very low (between 0 and 2 percent). Imputation was only performed for two surveys released before 2004; however, data have been imputed for all missing questionnaire data released thereafter. For the *Distance Education at Degree-Granting Postsecondary Institutions: 2006-07* survey, missing data were imputed using a "hot-deck" approach to obtain a "donor" institution from which the imputed values were derived. Under the hot-deck approach, a donor institution that matched selected characteristics of the institution with missing data (the recipient institution) was identified. Once a donor was found, it was used to derive the imputed values for the institution with missing data. For categorical items, the imputed value was simply the corresponding value from the donor institution. For numerical items, the imputed value was calculated by taking the donor's response for that item (e.g., enrollment in dual enrollment programs) and dividing that number by the total number of students enrolled in the donor institution. This ratio was then multiplied by the total number of students enrolled in the recipient institution to provide an imputed value. All missing items for a given institution were imputed from the same donor whenever possible.

Sampling Error

Estimates are based on the selected samples and, consequently, are subject to sampling variability. The standard error is a measure of the variability of estimates due to sampling. Because the data from PEQIS surveys are collected using a complex sampling design, the variances of the estimates from the surveys (e.g., estimates of proportions) are typically different from what would be expected from data collected with a simple random sample. To generate accurate standard errors for the estimates, standard errors are computed using a technique known as jackknife replication. The standard errors were calculated using a computer program.

Nonsampling Error

Nonsampling error describes variations in the estimates that may be caused by population coverage limitations and data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems like unit and item nonresponse, differences in respondents' interpretations of the meaning of questions, response differences related to the particular time the survey was conducted, and mistakes made during data preparation. It is difficult to identify and estimate either the amount of nonsampling error or the bias caused by this error. To minimize the potential for nonsampling error, the *Distance Education at Degree-Granting Postsecondary Institutions: 2006-07* survey used a variety of procedures, including a pretest of the questionnaire with the individual at each postsecondary institution deemed to be the most knowledgeable about its distance education programs and courses. The pretest provided the opportunity to check for consistency in the interpretation of questions and definitions and to eliminate ambiguous items. The questionnaire and instructions were also extensively reviewed by NCES and the data requestor at the Office of Educational Technology. In addition, both manual editing and machine editing of the questionnaire

responses were conducted to check the data for accuracy and consistency. Cases with missing or inconsistent items were recontacted by telephone to resolve problems. Data were keyed with 100 percent verification for surveys received by mail, fax, or telephone.

Coverage Error. Because the sampling frames for PEQIS surveys are constructed from IPEDS data files, coverage error is believed to be minimal.

Nonresponse Error. Both unit nonresponse and item nonresponse are quite low in PEQIS surveys. For the 16 surveys completed thus far, weighted unit response has ranged from 87 to 97 percent (see table 25). Item nonresponse for most items in PEQIS surveys has been less than 1 percent. The weights are adjusted for unit nonresponse.

For the PEQIS *Dual Enrollment of High School Students at Postsecondary Institutions: 2002-03*, 23 institutions were determined to be ineligible for the panel. For the eligible institutions, an unweighted response rate of 92 percent (1,460 responding institutions divided by the 1,590 eligible institutions in the sample for this survey) was obtained. The weighted response rate for this survey was 93 percent. The unweighted overall response rate was 91 percent (99 percent panel participation rate multiplied by the 92 percent survey response rate). The weighted overall response rate was 92 percent (99 percent weighted panel participation rate multiplied by the 93 percent weighted survey response rate).

Measurement Error. This type of nonsampling error may result from different interpretations of survey definitions by respondents or from the institution's inability to report according to survey specifications due to deficiencies in its recordkeeping system. Some examples of measurement error in PEQIS surveys follow.

For the PEQIS *Distance Education at Degree-Granting Postsecondary Institutions: 2006-07*, approximately 20 institutions were determined to be ineligible for the panel. For the eligible institutions, an unweighted response rate of 90 percent (1,450 responding institutions divided by the 1,610 eligible institutions in the sample for this survey) was obtained. The weighted response rate for this survey was 87 percent.

The 1995 Survey on Remedial Education in Higher Education Institutions was conducted to provide current national estimates on the extent of remediation on college campuses. Institutions provided information about the remedial reading, writing, and mathematics courses they offered in fall 1995. Remedial courses were defined as courses designed for college students lacking the skills necessary to perform college-level work at the level required by the institution. Thus, what constituted remedial courses varied by institution. Respondents were asked to include any courses meeting the definition, regardless of name. Some institutions refer to remedial courses as "compensatory," "developmental," or "basic skills."

Table 25. Weighted unit response rates for recent PEQIS surveys: Selected years, 2000–10

Survey	Panel participation rate	Weighted 1st level response rate	Overall weighted response rate
Distance Education at Postsecondary Institutions, 2006-07	†	87	87
Educational Technology in Teacher Education Programs for Initial Licensure	†	95	95
Dual Enrollment Programs and Courses for High School Students	99	93	92
Distance Education at Postsecondary Education Institutions, 2000-01	99	94	93

† Not applicable.
SOURCE: U.S. Department of Education, National Center for Education Statistics. (2009). Public-Use Data Files and Documentation (PEQIS 16): Distance Education at Postsecondary Institutions, 2006-07 (NCES 2009-074). U.S. Department of Education, National Center for Education Statistics. (2008). Educational Technology in Teacher Education Programs for Initial Licensure (PEQIS 15): Public-Use Data Files and Documentation (NCES 2008-013). U.S. Department of Education, National Center for Education Statistics. (2009). Public Use Data Files and Documentation (PEQIS 14): Dual Enrollment Programs and Courses for High School Students (NCES 2009-045). U.S. Department of Education, National Center for Education Statistics. (2005). Distance Education at Higher Education Institutions: 2000-01 (PEQIS 13): Public-Use Data Files and Documentation (NCES 2005-118).

Data Comparability

While most PEQIS surveys are not designed specifically for comparison with other surveys, the data from some PEQIS surveys can be compared with data from other postsecondary surveys. There have been, however, four administrations of the PEQIS *Survey on Distance Education Courses Offered by Higher Education Institutions* and two administrations of the *Survey on Remedial Education in Higher Education Institutions*.

The 1998 *Survey on Students With Disabilities at Postsecondary Education Institutions* complements another NCES study on the self-reported preparation, participation, and outcomes of students with disabilities. The latter study is based on an analysis of four different NCES surveys, which were used to address enrollment in postsecondary education, access to postsecondary education, persistence to degree attainment, and early labor market outcomes and graduate school enrollment rates of college graduates with disabilities. See *Students With Disabilities in Postsecondary Education: A Profile of Preparation, Participation, and Outcomes* (Horn and Berktold 1998).

The four administrations of the *Survey on Distance Education Courses Offered by Higher Education Institutions*—conducted first in late 1995, then in 1998–99, in 2000–01, and again in 2006–07—were the first to collect nationally representative data about distance education course offerings in higher education institutions. The four surveys differed in their samples and question wording. The sample for the first distance education survey, conducted in 1995, consisted of 2-year and 4-year higher education institutions in the 50 states, the District of Columbia, and Puerto Rico. At the time, NCES defined higher education institutions as institutions accredited at the college level by an agency recognized by the Secretary of the U.S. Department of Education. Higher education institutions were a subset of all postsecondary institutions. The sample for the second distance education survey, conducted in winter 1998–99, consisted of 2-year and 4-year postsecondary institutions (both higher education and other postsecondary institutions) in the 50 states and the District of Columbia. The third survey, conducted in 2000–01, included 2-year and 4-year Title IV-eligible, degree-granting institutions in the 50 states and the District of Columbia. Furthermore, data from the 1995 and 2000–01 surveys were not imputed for item nonresponse; however, comparisons between the surveys are possible when using the subset of higher education institutions from the 1998–99 survey. The fourth survey, conducted in 2006–07, also included 2-year and 4-year Title IV-eligible, degree-granting institutions in the 50 states and the District of Columbia. While this survey covered many of the same topics covered in the previous surveys, the data are not comparable because the definition of distance education in the 2006–07 survey reflected two major changes: First, the definition no longer included a criterion for instructional delivery to off-campus or remote locations; second, the definition included correspondence courses and distance education courses that were designated by institutions as hybrid/blended online courses.

The 1995 and 2000 administrations of the *Survey on Remedial Education in Higher Education Institutions* were conducted to provide national estimates on the extent of remediation on college campuses. The results update the information collected in two earlier NCES surveys for academic years 1983–84 and 1989–90; because PEQIS was not yet in existence, these surveys were conducted under the FRSS. (See section 1 of this chapter.) Although the 1995 survey was not designed as a comparative study, the results can be compared with data from the 1993–94 IPEDS Institutional Characteristics Survey: PEQIS estimated that 78 percent of institutions offered at least one remedial course for freshmen in fall 1995, and IPEDS estimated that 79 percent of institutions offered remedial courses in academic year 1993–94. At the student level, results from the 1995 PEQIS survey can be compared with results from institutional surveys conducted by the American Council on Education as well as a study conducted by the Southern Regional Education Board. However, these studies asked about freshmen needing remediation rather than about freshmen enrolled in remedial courses.

The remedial education data from the 1989 and 1995 surveys are not comparable to the data from the 2000 survey because of a change in the way that NCES categorized postsecondary institutions (and because of the inclusion of institutions in Puerto Rico in the earlier surveys). The data for the 1989 and 1995 surveys represent 2-year and 4-year higher education institutions that enroll freshmen. At the time these surveys were conducted, NCES defined higher education institutions as institutions accredited at the college level by an agency recognized by the Secretary of the U.S. Department of Education. Higher education institutions were a subset of all postsecondary institutions. The data for the 2000 survey represent 2-year and 4-year Title IV-eligible, degree-granting institutions that enroll freshmen. This change was necessary because the Department of Education stopped making a distinction between higher education institutions and other postsecondary institutions eligible to participate in federal Title IV financial aid

programs; thus, NCES no longer categorized institutions as higher education institutions. In order to make comparisons between the 1995 and 2000 surveys, the data from the 1995 survey can be reanalyzed with the definition of eligible institutions changed to match the definition for the 2000 survey as closely as possible.

Remedial enrollment can also be examined using postsecondary transcripts collected from institutions during the National Longitudinal Study of the High School Class of 1972 and the High School and Beyond Longitudinal Study (See chapters 6 and 7), as well as from student reported data in National Postsecondary Student Aid Study (NPSAS) (see chapter 14). Institutional reports of remedial enrollment in all of these surveys are substantially higher than student self-reports collected in NPSAS.

Contact Information

For content information on PEQIS, contact:

Peter C. Tice
Phone: (202) 502-7497
E-mail: peter.tice@ed.gov

Mailing Address:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

Methodology and Evaluation Reports

Methodology is discussed in the technical notes to survey reports. Some of these reports are listed below.

Greene, B. (2005). *Distance Education at Higher Education Institutions: 2000–01* (NCES 2005-118). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Horn, L., and Berktold, J. (1998). *Students With Disabilities in Postsecondary Education: A Profile of Preparation, Participation, and Outcomes* (NCES 1999-187). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Kleiner, B., and Lewis, L. (2005). *Dual Enrollment of High School Students at Postsecondary Institutions: 2002–03* (NCES 2005-008). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kleiner, B., Thomas, N., and Lewis, L. (2007). *Educational Technology in Teacher Education Programs for Initial Licensure* (NCES 2008-040). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Lewis, L., Alexander, D., and Farris, E. (1998). *Distance Education in Higher Education Institutions* (NCES 98-062). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Lewis, L., and Farris, E. (1997). *Campus Crime and Security at Postsecondary Education Institutions* (NCES 97-402). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Lewis, L., and Farris, E. (1999). *An Institutional Perspective on Students With Disabilities in Postsecondary Education* (NCES 1999-046). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Lewis, L., Snow, K., Farris, E., and Levin, D. (2000). *Distance Education at Postsecondary Education Institutions: 1997–98* (NCES 2000-013). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Parsad, B., and Lewis, L. (2003). *Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000* (NCES 2004-010). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Parsad, B., and Lewis, L. (2008). *Distance Education at Degree-Granting Postsecondary Institutions: 2006–07* (NCES 2009-044). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Phelps, R., Parsad, B., Farris, E., and Hudson, L. (2001). *Features of Occupational Programs at the Secondary and Postsecondary Education Levels* (NCES 2001-018). National Center for Education Statistics, U.S. Department of Education. Washington, DC.

Appendix A: Glossary of Statistical Terms

Balanced Incomplete Block (BIB) spiraling: In a BIB design, as in standard matrix sampling, no sample unit is administered all of the tasks in the assessment pool. However, unlike standard matrix sampling (in which items or tasks are assembled into discrete booklets), BIB design requires that sample units receive different interlocking sections of the assessment forms that allow for the estimation of relationships among all the tasks in the pool through the unique linking of blocks.

“Spiraling” refers to the method by which test booklets are assigned to sample units. Each version of the assessment booklet must appear in the sample approximately the same number of times and must be administered to equivalent subgroups within the full sample. To ensure proper distribution at assessment time, the booklets are packed in spiral order (e.g., one each of booklets 1 through 7, then 1 through 7 again, and so on). The test coordinator randomly assigns these booklets to the sample units in each test administration session. Spiraled distribution of the booklets promotes comparable sample sizes for each version of the booklet, ensures that these samples are randomly equivalent, and reduces the likelihood that sample units will be seated within viewing distance of an identical booklet.

Balanced Repeated Replication (BRR): See Replication technique.

Base weight: The product of the reciprocals of the probabilities of inclusion for all stages of sampling.

Bias (due to nonresponse): The difference that occurs when respondents differ as a group from nonrespondents on a characteristic being studied.

Bias (of an estimate): The difference between the expected value of a sample estimate and the corresponding true value for the population.

Blanking edit: See Edit.

Bootstrap: See Replication techniques.

CAPI: Computer Assisted Personal Interviewing enables data collection staff to use portable microcomputers to administer a data collection form while viewing the form on the computer screen. As responses are entered directly into the computer, they are used to guide the interview and are automatically checked for specified range, format, and consistency.

Chi-squared Automatic Interaction Detector (CHAID) Analysis: This technique divides the respondent data into segments which differ with respect to the item being imputed. This segmentation process first divides the data into groups based on categories of the most significant predictors. It then splits each of these groups into smaller groups based on other predictor variables and merges categories of a variable found insignificant (by x^2 test). This splitting and merging progress continues until no more statistically significant predictors are found. The imputation classes form the final CHAID segments.

Cohort: A group of individuals who have a statistical factor in common (e.g., year of birth, grade in school, year of high school graduation).

“Cold-deck” imputation: See Imputation.

Component weight: For each stage of sampling, the component weight is equal to the reciprocal of the probability of selecting the unit at that stage.

Composite variable: Variable constructed through the combination of two or more variables (e.g., socioeconomic status) or through calculation by applying a mathematical function to a variable. Composite variable is also referred to as derived, constructed, or classification variable.

Computer Assisted Personal Interviewing: See CAPI.

Consistency edit: See Edit.

Coverage error: Coverage error in an estimate results from the omission of part of the target population (undercoverage) or the inclusion of units from outside the target population (overcoverage).

Critical items or key items: Items deemed crucial to the methodological or analytical objectives of the study.

Design effect: The cumulative effect of the various design factors affecting the precision of statistics is often modeled as the sample design effect. The design effect, DEFF, is defined as the ratio of the sampling variance of the statistic (e.g., a mean or a proportion) under the actual sampling design divided by the variance that would be expected for a simple random sample of the same size. Hence, the design effect is equal to one, by definition, for simple random samples. For clustered multistage sampling designs, the design effect is greater than unity, reflecting that the precision is less than could be achieved with a simple random sample of the same size (if that were the sampling design). The size of the design effect depends largely on the intracluster correlation of the survey observations within the primary sampling units. Hence, statistics that are based on observations that are highly correlated within units will have higher design effects.

Durbin's Method: This method selects two first-stage units per stratum without replacement, with probability proportional to size so that the joint inclusion probabilities are greater than zero for every pair.

Edits: Procedures for checking and modifying response in a survey.

> ***Blanking edit:*** Deletes extraneous entries and assigns the "not answered" code to items that should have been answered but were not.
>
> ***Consistency edit:*** Identifies inconsistent entries within each record and, whenever possible, corrects them. If they cannot be corrected, the entries are deleted. Inconsistencies can be (1) within items or (2) between items. The consistency edit also fills some items where data are missing or incomplete by using other information on the data record.
>
> ***Logic edit:*** Checks made of the data to ensure logical consistency among the responses from a data provider.
>
> ***Range check:*** Determines whether responses fall within a predetermined set of acceptable values.
>
> ***Relational edit check:*** Compares data entries from one section of the questionnaire for consistency with data entries from another section of the questionnaire.
>
> ***Skip pattern check:*** Checks if responses correctly followed skip pattern instructions.
>
> ***Summation check:*** Compares reported totals with the sums of the constituent data items.

Estimate: A numerical value obtained from a statistical sample and assigned to a population parameter. The particular value yielded by an estimator in a given set of circumstances or the rule by which such particular values are calculated.

Estimation: Estimation is concerned with inference about the numerical value of unknown population values from incomplete data, such as a sample. If a single figure is calculated for each unknown parameter, the process is called point estimation. If an interval is calculated within which the parameter is likely, in some sense, to lie, the process is called interval estimation.
Field test: The study of a data collection activity in the setting where it is to be conducted.

“Hot-deck” imputation: See Imputation.

Imputation: Imputation (for item or survey nonresponse) involves supplying a value if an item response is missing. The items may be missing because the respondent was careless, refused to provide an answer, or could not obtain the requested information. Since extensive amounts of missing data can seriously bias sample-based estimates, procedures for imputing missing values are often developed. Imputation is used to reduce nonresponse bias in survey estimates, simplify analyses, and improve the consistency of results across analyses.

Depending on the type of data to be imputed and the extent of missing values, a number of alternative techniques can be employed. These techniques include: logical imputation, the use of poststratum averages, “hot deck” imputation, and regression and other “modeling” techniques.

> ***“Cold-deck” imputation:*** A process that imputes missing data with values observed from a past survey.
>
> ***“Hot-deck” imputation:*** Hot deck refers to a general class of procedures for which cases with missing items are assigned the corresponding value of a “similar” respondent in the sample.
>
> > *Random within class:* This method divides the total sample into imputation classes according to the values of the auxiliary variables. Each nonrespondent is assigned a value randomly selected from the same imputation class.
> >
> > *Sequential (also known as traditional):* The records of the survey are treated sequentially in the same imputation class and for each class a single value is stored to provide a starting point for a single pass through the data file. If a record has a response, that value replaces the previous value. If the record is missing, the currently stored value is assigned to that unit.
>
> ***Logical imputation:*** Logical imputation can be applied in situations where a missing response can be inferred with certainty (or high degree of probability) from other information in the data record.
>
> ***Poststratum average:*** In the use of poststratum average, a record with missing data is assigned the mean value of those cases in the same “poststratum” for which information on the item is available.
>
> ***Proc Impute:*** This is an advanced software package that performs three steps for each target variable to be imputed:
>
> > 1) Uses stepwise regression analysis to find the best combination of predictors among all variables included in the imputation model;
> >
> > 2) Creates homogeneous cells of records which have close predicted regression values; and
> >
> > 3) Imputes each missing record in a given cell with a weighted average of two donors, one from its own cell and the other from an adjacent cell.
>
> ***Regression and other modeling techniques:*** These techniques operate by modeling the variable to be imputed, Y as a function of related independent variables, $X_1, X_2, \ldots, X_p$. To preserve the variability of the Y’s at specific values of $X_1, X_2, \ldots, X_p$, a residual, $\hat{e}$, is sometimes added to the predicted value determined from the model.

Independent variable: In regression analysis, when a random variable, Y, is expressed as a function of variables $X_1, X_2, \ldots, X_p$, plus a stochastic term, the X’s are known as “independent variables.”

Item nonresponse: An item on a data collection form that is missing when a response was expected.

Jackknife method: See Replication techniques.
Key item or critical item: Item deemed crucial to the methodological or analytical objectives of the study.

Keyfitz approach: A method of probability selection that maximizes the selected units from a past sample.

Logic edit: See Edit.

Logical imputation: See Imputation.

Measurement error: Measurement error refers to errors in estimates resulting from incorrect responses gathered during the data collection phase of a survey. Measurement errors result, for instance, when the respondent gives (intentionally or unintentionally) incorrect answers, the interviewer misunderstands or records answers incorrectly, the interviewer influences the responses, the questionnaire is misinterpreted, etc.

Mitofsky-Waksberg method: A method of sample selection for household telephone interviewing via random digit dialing where the sampling is carried out through a two-stage design. As Waksberg explained in his 1978 *Journal of the American Statistical Association* article: "Obtain from AT&T a recent list of all telephone area codes and existing prefix numbers within the areas. To these add all possible choices for the next two digits, and thus prepare a list of all possible first eight digits of the ten digits in telephone numbers. These eight-digit numbers are treated as Primary Sampling Units (PSU). A random selection is then made of an eight-digit number, and also of the next two digits. The number is then dialed. If the dialed number is at a residential address, the PSU is retained in the sample. Additional last two digits are selected at random and dialed within the same eight-digit group, until a set number, k, of residential telephones is reached. Interviews are attempted both at the initial number and the additional k numbers. If the original number called was not residential, the PSU is rejected. This process is repeated until a predesignated number of PSU's, m, is chosen. The total sample size is, therefore, $m(k + 1)$. The values of m and k are chosen to satisfy criteria for an optimum sample design." Note that although all units have the same probabilities of selection, it is not necessary to know the probabilities of selection of the first-stage or the second-stage units.

Nonresponse: Cases in data collection activities in which potential data providers are contacted but refuse to reply or are unable to do so for reasons such as deafness or illness.

Nonresponse bias: This occurs when respondents as a group differ from nonrespondents in their answers to questions on a data collection form.

Nonsampling error: This term is used to describe variations in the estimates that may be caused by population coverage limitations, as well as data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems like unit and item nonresponse, the differences in respondents' interpretations of the meaning of the questions, response differences related to the particular time the survey was conducted, and mistakes in data preparation.

Open-ended: A type of interview question that does not limit the potential response to predetermined alternatives.

Out-of-range response: A response that is outside of the predetermined range of values considered acceptable for a particular item.

Oversampling: Deliberately sampling a portion of the population at a higher rate than the remainder of the population.

Plausible value: Proficiency value drawn at random from a conditional distribution of a survey respondent, given his or her response to cognitive exercises and a specified subset of background variables.

Plausible values methodology: Plausible values methodology represents what the true performance of an individual might have been, had it been observed, using a small number of random draws from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background

variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. The draws from the distribution are different from one another to quantify the degree of precision (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performances.

Population: All individuals in the group to which conclusions from a data collection activity are to be applied.

Poststratification: An estimation method that adjusts the sampling weights so that they sum to specified population totals corresponding to the levels of a particular response variable.

Poststratification adjustment: A weight adjustment that forces survey estimates to match independent population totals within selected poststrata (adjustment cells).

Precision: The difference between a sample-based estimate and its expected value. Precision is measured by the sampling error (or standard error) of an estimate.

Pretest: A test to determine performance prior to the administration of a data collection activity.

Probability sample: A sample selected by a method such that each unit has a fixed and determined probability of selection.

Proc Impute: See Imputation.

Processing: The manipulation of data.

Range check: See Edit.

Regression analysis: A statistical technique for investigating and modeling the relationship between variables.

Relational edit check: See Edit.

Replicate estimate: An estimate of the population quantity based on the replicate subsample using the same estimation methods used to compute the full sample estimate.

Replicate sample: One of a set of subsamples, each obtained by deleting a number of observations in the original sample for the purpose of computing the appropriate variance based on the complex design of the survey.

Replicate weight: The weight assigned to an observation for a particular replicate subsample.

Replicate: A term often used to refer to either the replicate sample or the replicate estimate, depending on context.

Replication technique: Method of estimating sampling errors that involve repeated estimation of the same statistic using various subsets of data providers. The major methods are balanced repeated replication (BRR), bootstrap, and the jackknife technique.

> ***Balanced Repeated Replication (BRR):*** A method of replication that divides the sample into half-samples.
>
> ***Bootstrap:*** A resampling technique of creating replicates by drawing random samples with replacement that mirror the original sampling plan for a pseudo-population constructed from the original sample.
>
> ***Jackknife method:*** A method of replication that creates replicates (subsets) by excluding one unit at a time from the sample.

Sample: A subgroup selected from the entire population.

Sampling error: When a sample rather than the entire population is surveyed, estimates can differ from the true population values that they represent. This difference, or sampling error, occurs by chance, and its variability is measured by the standard error of the estimate. Sample estimates from a given survey design are unbiased when an average of the estimates from all possible samples would yield, hypothetically, the true population value. In this case, the sample estimate and its standard error can be used to construct approximate confidence intervals, or ranges of values, that include the true population value with known probabilities.

Sampling variance: A measure of dispersion of values of a statistic that would occur if the survey were repeated a large number of times using the same sample design, instrument, and data collection methodology. The square root of the sampling variance is the standard error.

Sampling weight: See Weighted estimate.

Scaling (item response theory): Item response theory (IRT) scaling assumes some uniformity in response patterns when items require similar skills. Such uniformity can be used to characterize both examinees and items in terms of a common scale attached to the skills, even when all examinees do not take identical sets of items. Comparisons of items and examinees can then be made in reference to a scale, rather than to the percent correct. IRT scaling also allows the distributions of examinee groups to be compared.

This is accomplished by modeling the probability of answering a question in a certain ways as a mathematical function of proficiency or skill. The underlying principle of IRT is that, when a number of items require similar skills, the regularities observed across patterns of response can often be used to characterize both respondents and tasks in terms of a relatively small number of variables. When aggregated through appropriate mathematical formulas, these variables capture the dominant features of the data. IRT enables the assessment of a sample of students in a subject area or subarea on a common scale even if different students have been administered different exercises.

Skip pattern check: See Edit.

Special population: A subset of the total population distinguishable by unique needs, characteristics, or interests (e.g., disadvantaged students, gifted students, physically or mentally handicapped students, vocational education students).

Spiraling: See Balanced Incomplete Block (BIB) spiraling.

Standard deviation: The most widely used measure of dispersion of a set of values. It is equal to the positive square root of the population variance.

Standard error: The positive square root of the sampling variance. It is a measure of the dispersion of the sampling distribution of a statistic. Standard errors are used to establish confidence intervals for the statistics being analyzed.

Statistically significant: There is a low probability that the result is attributable to chance alone.

Summation check: See Edit.

Taylor Series: The Taylor Series variance estimation procedure is a technique to estimate the variances of nonlinear statistics. The procedure takes the first-order Taylor Series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. For stratified multistage surveys such as the National Postsecondary Student Aid Study (NPSAS), the Taylor Series procedure requires analysis strata and analysis replicates defined from the sampling strata and primary sampling units (PSUs) used in the first stage of sampling.

Target population: See Population.

Time series: A set of ordered observations on a quantitative characteristic of an individual or collective phenomenon taken at different points in time. Usually the observations are successive and equally spaced in time.

Unit nonresponse: The failure of a survey respondent to provide any information.

Variable: A quantity that may assume any one of a set of values.

Variance: See Population variance and Sampling variance.

Weighted estimate: Estimate from a sample survey in which the sample data are weighted (multiplied) by factors reflecting the sample design. The weights (referred to as sampling weights) are typically equal to the reciprocals of the overall selection probabilities, multiplied by a nonresponse or poststratification adjustment.

Appendix B: Ordering NCES Publications and Data Files

Much NCES data and many NCES publications are available through the NCES web site. The NCES Electronic Catalog ((http://nces.ed.gov/pubsearch/) allows searching for NCES products by NCES number or, for products released within the last 5 years, by keyword, survey/program area, type of product, and release date. The Electronic Catalog also has lists of publications published in the last 90 days, data products released in the last 6 months, and all publications and data products by survey and program area.

In addition to downloading from the NCES web site, there are four other ways to obtain NCES publications, CD-ROMs, and other products:

1. Education Publications Center (ED Pubs),
2. Government Printing Office (GPO),
3. Federal Depository Libraries, and
4. Education Resources Information Center (ERIC).

Education Publications Center (ED Pubs)

Until supplies are exhausted, a single copy of a publication or CD-ROM may be obtained at no cost from ED Pubs. Before requesting a copy, it is necessary to have the complete title and NCES number for the publication (e.g., The Condition of Education, 2010, NCES 2010–028).

Toll-free number: (877) 4ED–PUBS, (877) 433–7827
TTY/TDD toll-free number: (877) 576–7734
Para español, llame al: (877) 433–7827 (toll-free)
Fax: (703) 605–6794
E-mail: edpubs@edpubs.ed.gov
Internet: www.edpubs.org/

Mailing Address:

ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

Government Printing Office (GPO)

If more than one copy of a publication is needed, or if ED Pubs' supplies have been exhausted, many—not all—NCES products may be purchased from the Government Printing Office (GPO). To order a copy from GPO, it is necessary to have the product's GPO stock number (e.g., 065-000-00871-8). The product's stock number and price can be found out by going to the U.S. Government Online Bookstore and entering the product's title or keyword.

Online orders: http://bookstore.gpo.gov/
Phone orders: 1–866–512–1800 (toll-free); (202) 512–1800 (DC area)
Fax: Credit card orders may be faxed to (202) 512–2104
Email: contactcenter@gpo.gov

Mailing Address:
U.S. Government Printing Office
Mail Stop: IDCC
732 N. Capitol Street, NW
Washington, DC 20401

Federal Depository Library
For older publications, the only source for an NCES publication may be a Federal depository library. There are nearly 1,250 of these libraries around the country. However, only the "Regional" libraries receive all materials distributed through the Federal Depository Library Program. Other Federal depository libraries select materials according to the needs of their communities. These libraries can be located through the following web site:

http://www.gpoaccess.gov/libraries.html

Education Resources Information Center (ERIC)
ERIC is an online digital library of education research and information. ERIC is sponsored by the Institute of Education Sciences (IES) of the U.S. Department of Education. ERIC provides access to education literature to support the use of educational research and information to improve practices in learning, teaching, educational decision-making, and research.

Toll-free number: (800) LET-ERIC, (800) 538-3742
Internet: http://www.eric.ed.gov/

Mailing Address:
ERIC Project
c/o Computer Science Corporation
655 15th Street, NW
Washington, DC 20005

Public-use versus Restricted-use Data Files
NCES uses the term "public-use data" for survey data when the individually identifiable information has been coded or deleted to protect the confidentiality of survey respondents. All NCES public-use data files can be accessed (at no cost) from the NCES web site. Only public-use data files that are on CD-ROM are available through ED Pubs or GPO.

Restricted-use data files contain individually identifiable information, which is confidential and protected by law. To use these data, researchers must obtain a restricted-use data license. A brief summary of the steps that need to be followed to obtain (or amend) a restricted-use data license is provided below. The procedures are fully discussed in the *NCES Restricted-Use Data Procedures Manual*, which can be found at the following web site:

http://nces.ed.gov/statprog/rudman/

As of July 1, 2007, NCES requires restricted-use data license applications and amendments to be submitted through its new Electronic Application System, accessible at the following web site:

http://nces.ed.gov/statprog/instruct.asp

The following information will be required to obtain or amend a restricted data license. For more detailed information, applicants should see the Restricted-Use Data Procedures Manual.

1. The license number to be amended (if the researcher already has a license);

2. The name of the dataset(s) the researcher wishes to use;

3. The purpose for the loan of the data;

4. The length of time the researcher will need the data (loan period not to exceed five years);

5. The computer security plan the researcher will follow;

6. The list of authorized users; and

7. An affidavit of nondisclosure for each authorized user, promising to keep the data completely confidential.

A researcher who is amending an existing license and whose purpose is a continuation of the project that was approved originally may be able to condense the abstract of the research design, but the description must be specific enough to justify using the raw data. Similarly, researchers who plan to use the same computer(s) and person(s) who are already licensed users may be able to simply update the computer security plan previously approved. Computer security plans need to be followed carefully as spot site inspections do occur. In the case of postsecondary institutions, only faculty can serve as the primary project officer overseeing the daily operations. Graduate students may be listed as authorized users only.

Mailing Address:
IES Data Security Office
Department of Education/IES/NCES
1990 K Street NW Room 9060

Washington, DC 20006-5574

Contact Person:
Cynthia L. Barton
Data Security Assistant
Phone: (202) 502–7307
E-mail: cynthia.barton@ed.gov

Note on Working Papers: Working papers are available on the NCES web site through the Electronic Catalog.

Appendix C: Web-based and Standalone Tools for Use with NCES Survey Data

NCES has developed a number of web-based and standalone tools for use with its data*. There are four user tools that have been developed for use across multiple surveys: the Data Analysis System (DAS), which produces tabular data for the user; Electronic Codebooks (ECBs), which allow users to develop datafiles in SAS, SPSS, or ASCII format; the Education Data Analysis Tool (EDAT), which allows the user to download NCES survey datasets; and the International Data Explorer (IDE), which is an interactive tool used to explore international assessment results. These are described in more detail below, along with a list of the surveys available with each. Following this, descriptions of the tools developed for more specialized uses—for example, the Private School Locator and the NAEP Test Questions Tool—are provided in a survey-by-survey list.

The Data Analysis System (DAS) (http://nces.ed.gov/das/) is a software application that provides access to Department of Education survey data. DAS allows users to create programming instruction files (DAS files) that specify the information they want displayed in a table. The output table will contain the estimates (usually percentages of students) and corresponding standard errors which are calculated taking into account the complex sampling designs used in NCES surveys. In addition, the DAS software can create correlation matrices which can be used as input for most popular statistical programs to conduct multivariate analysis. There is a separate DAS for each survey data set, and all have a consistent interface and command structure. DAS applications are available in Windows- and web-based formats. The available NCES surveys are:

- Baccalaureate and Beyond (B&B) Longitudinal Study
- Beginning Postsecondary Students (BPS) Longitudinal Study
- Early Childhood Longitudinal Study (ECLS)
- High School and Beyond (HS&B) Longitudinal Study
- Integrated Postsecondary Education Data System (IPEDS)
- National Education Longitudinal Study of 1988 (NELS:88)
- National Household Education Surveys (NHES) Program
- National Longitudinal Study of the High School Class of 1972 (NLS:72)
- National Postsecondary Student Aid Study (NPSAS)
- National Study of Postsecondary Faculty (NSOPF)

Electronic Codebook (ECB) programs have been created for many NCES surveys. These programs, after being installed on a user's personal computer, allow the user to examine the variables in each of a survey's data files, as

*As explained in appendix B, all NCES public-use data files can be accessed (at no cost) from the NCES web site. To use restricted-use data, researchers must first obtain a restricted-use data license.

well as create SAS and SPSS programs that will generate an extract data file from any of the survey data files on the CD-ROM.

ECB programs are usually included on a CD-ROM with the survey data, but NCES has issued a CD-ROM that contains only electronic codebooks. This CD-ROM was created to provide updated ECB software for data sets that were, in some cases, released several years ago.

ECBs may be available for use with public-use data, restricted-use data, or both, depending on the survey. ECBs are available for the following surveys:

- Baccalaureate and Beyond (B&B) Longitudinal Study—Restricted-use
- Beginning Postsecondary Students (BPS) Longitudinal Survey—Restricted-use
- Early Childhood Longitudinal Study (ECLS)—Public-use and Restricted-use
- Education Longitudinal Study of 2002 (ELS:2002) — Public-use and Restricted-use
- High School and Beyond (HS&B) Longitudinal Study—Restricted-use
- High School Transcript (HST) Study—Restricted-use
- Integrated Postsecondary Education Data System (IPEDS)—Public-use
- National Assessment of Adult Literacy (NAAL)— Public-use and Restricted-use
- National Education Longitudinal Study of 1988 (NELS:88)—Public-use and Restricted-use
- National Household Education Surveys (NHES) Program—Public-use and Restricted-use
- National Longitudinal Study of the High School Class of 1972 (NLS:72)—Public-use
- National Postsecondary Student Aid Study (NPSAS)—Restricted-use
- National Study of Postsecondary Faculty (NSOPF)—Public-use and Restricted-use
- Private School Universe Survey (PSS)—Public-use
- Progress in International Reading Literacy Study (PIRLS)—Public-use and Restricted-use
- Schools and Staffing Survey (SASS)—Public-use and Restricted-use
- Trends in International Mathematics and Science Study (TIMSS)—Public-use

The Education Data Analysis Tool (EDAT) (http://nces.ed.gov/edat/) is a web tool that allows the user to download NCES survey datasets. EDAT guides the user through 1) selecting a survey, 2) selecting variables relevant to the user's analysis, 3) downloading a data set to the user's computer, and 4) downloading syntax files. Statistical software package (SAS, SPSS, Stata, R, S-Plus, or SUDAAN) or a generic file format (ASCII or CSV) can be selected for download format. If a statistical software package is selected, EDAT will use generate a custom syntax file for use with the selected software. If a generic file format is chosen, EDAT will generate a layout file to help the user to with the data. In either case, EDAT will generate a codebook file with codes, labels, descriptions, and frequencies for the user's reference. There is a tutorial for this tool. The following surveys are currently accessible in EDAT (additional datasets will be added in the near future):

- Education Longitudinal Study of 2002 (ELS:2002)

The International Data Explorer (IDE) (http://nces.ed.gov/surveys/international/ide/) is a dynamic, interactive tool used to explore international assessment results for various subjects, grades, and jurisdictions. It allows users to create custom statistical tables, graphics, and maps using data from international assessments. There is a tutorial for this tool. Student performance in the context of gender, race/ethnicity, and many other factors can be examined using data gathered from participants in the following international assessments:

- Program for International Student Assessment (PISA)
- Progress in International Reading Literacy Study (PIRLS)
- Trends in International Mathematics and Science Study (TIMSS)

EARLY CHILDHOOD EDUCATION SURVEY

Early Childhood Longitudinal Study (ECLS)

- **DAS for ECLS**—See DAS, above.
- **ECB for ECLS**—Public-use and Restricted-use

ELEMENTARY AND SECONDARY EDUCATION SURVEYS

Common Core of Data (CCD)

- **CCD CD-ROM Interface**: After selecting one of the three databases—School, Agency, or State—the user enters search criteria in specific fields, in order to limit the number of records for review to a select group. These records (matching the search criteria) can be displayed in summary or detail format, and can be printed. Specific fields for the selected records may be chosen and data exported to be used with other software packages. There are a number of export formats available that can be used with spreadsheets, databases, word processing packages, and statistical software packages.
- **Build a Table** (http://nces.ed.gov/ccd/bat/): This application enables users to create customized tables of CCD public school data for states, counties, MSA's, districts and schools using data from multiple years.
- **Public School District Longitudinal Data Tool** (http://nces.ed.gov/edfin/longitudinal/index.asp): Use this tool to compare fiscal and nonfiscal school district data over time from 1990 to 2002.
- **National Public School and School District Locator** (http://nces.ed.gov/ccd/search.asp): The School/District Locator enables users to find the correct name, address, telephone number, NCES ID number, locale (rural, large city, etc.), and other student and teacher information for public schools or school districts for the latest school year as reported to NCES by state education officials in each state. The Locator includes a Locator Glossary, which includes variable codes and definition descriptions, and a list of newly reported schools and school districts (this information is from unedited state data submissions and is updated as new information is received).

Search For Public Schools (http://nces.ed.gov/ccd/schoolsearch/): Use the Search For Public Schools locator to retrieve information on public schools from CCD's databases.

Search For Public School Districts (http://nces.ed.gov/ccd/districtsearch/): Use the Search For Public School Districts locator to retrieve information on public school districts from CCD's databases.

- **Public School District Finance Peer Search** (http://nces.ed.gov/edfin/search/search_intro.asp): This search allows users to compare the finances of a school district with its peers (those districts which share similar characteristics to the one chosen). Users may enter the entire name or only a portion of it. If more than one district with that name is found, users are prompted to select the appropriate one. Once the user has narrowed the search to one district, peer districts will be selected based on: enrollment, student/teacher ratio, median household income, district type, and metro status location. Users can base

their search for peers on a different set of criteria using the "Advanced" feature. Users wishing to perform a search other than a peer search may use the "Expert" feature.

- **State Education Data Profiles** (http://nces.ed.gov/programs/stateprofiles/): Use the state profiles tool to search for statewide information on elementary/secondary education characteristics and finance, postsecondary education, public libraries, assessments, and selected demographics for all states.

Education Longitudinal Study of 2002 (ELS:2002)

- **ECB for ELS:2002**— Public-use and Restricted-use

- **EDAT for ELS:2002**—See EDAT, above.

High School and Beyond (HS&B) Longitudinal Study

- **DAS for HS&B**—See DAS, above.

- **ECB for HS&B**—Restricted-use

National Education Longitudinal Study of 1988 (NELS:88)

- **DAS for NELS:88**—See DAS, above.

- **ECB for NELS:88**—Public-use and Restricted-use

National Longitudinal Study of the High School Class of 1972 (NLS-72)

- **DAS for NLS-72**—See DAS, above.

- **ECB for NLS-72**—Public-use

Private School Universe Survey (PSS)

- **ECB for PSS**—Public-use

- **Private School Locator** (http://nces.ed.gov/surveys/pss/privateschoolsearch/): The Private School Locator enables users to find school names, addresses, and other school information for private schools. Information for a particular private school, or a group of private schools, can be retrieved based on selection criteria the user specifies. Users can also download an ASCII text data file of the schools selected once the selection process is completed. The information in this locator comes from the approximately 30,000 schools that participated in the latest NCES Private School Survey (PSS). Users can request that schools not found in the Locator be included in future PSS. The Locator is also available on CD-ROM.

Schools and Staffing Survey (SASS)

- **ECB for SASS**—Public-use and Restricted-use

- **Table library** (http://nces.ed.gov/surveys/sass/tables.asp): The table library allows users to locate cross-tabulated data (and associated standard errors) produced from the Schools and Staffing Survey (SASS). These data tables came from the state and affiliation tables corresponding to the tables in NCES 2006-313, or the most frequently requested unpublished data, as well as other tables that had been produced but were not part of published reports. Additional tables and years of data will be included as they become available.

LIBRARY SURVEYS

Academic Library Survey (ALS)

- **Academic Library Peer Comparison Tool** (http://nces.ed.gov/surveys/libraries/compare/index.asp?LibraryType=Academic): This tool allows the user to get information on a particular library, or to customize a peer group by selecting the key variables that are used to define it. The user can then view customized reports of the comparison between the library of interest and its peers, on a variety of variables as selected by the user.

Public Libraries Survey (PLS)

- **Public Library Locator** (http://nces.ed.gov/surveys/libraries/librarysearch/): This tool helps users locate information about a public library or a public library service outlet when users know some, but not all of the information about it. The information in this locator has been drawn from the NCES Public Libraries Survey.

POSTSECONDARY AND ADULT EDUCATION SURVEYS

Baccalaureate and Beyond (B&B) Longitudinal Study

- **DAS for B&B**—See DAS, above.

- **ECB for B&B**—Restricted-use

Beginning Postsecondary Students (BPS) Longitudinal Study

- **DAS for BPS**—See DAS, above.

- **ECB for BPS**—Restricted-use

Integrated Postsecondary Education Data System (IPEDS)

- **ECB for IPEDS**—Public-use

- **IPEDS College Navigator** (http://nces.ed.gov/collegenavigator/): This is a direct link to over 9,000 colleges and universities in the United States. It was developed after NCES was authorized by Congress in 1998 to help college students, future students, and their parents understand the differences between colleges and how much it costs to attend college. Users can name a specific college or set of colleges and obtain information about them or use the search feature to find a college based on its location, program, or degree offerings either alone or in combination. The more criteria the user specifies, the smaller the number of colleges that will fit the criteria. Once the user has identified some colleges of interest, he or she can obtain important and understandable information on all of them.

- **IPEDS Peer Analysis System (Postsecondary Institutions)** (http://nces.ed.gov/ipedspas/): This tool is designed to enable a user to easily compare a postsecondary institution of the user's choice to a group of peer institutions, also selected by the user. This is done by generating reports using selected IPEDS variables of interest. There are tutorials for this tool.

National Postsecondary Student Aid Study (NPSAS)

- **DAS for NPSAS**—See DAS, above.

- **ECB for NPSAS**—Restricted-use

National Study of Postsecondary Faculty (NSOPF)

- **DAS for NSOPF**—See DAS, above.

ECB for NSOPF—Public-use and Restricted-use

EDUCATIONAL ASSESSMENT SURVEYS

National Assessment of Adult Literacy (NAAL)

- **ECB for NAAL**—Public-use and Restricted-use

National Assessment of Educational Progress (NAEP)

- **NAEP Data Tool Kit, including NAEPEX**: This is a data extraction program for choosing variables, extracting data, and generating SAS and SPSS control statements, and analysis modules for cross-tabulation and regression that work with SPSS and Excel (available on CD-ROM).

- **NAEP Test Questions Tool** (http://nces.ed.gov/nationsreportcard/itmrls/): The purpose of this tool is to provide easy access to NAEP questions, actual student responses, and scoring guides used in released portions of the NAEP assessments. National and, where appropriate, state data are also presented. Note that entire NAEP assessments are not presented here, since some questions must be kept secure for use in future NAEP assessments. Science is currently available only as a PDF document. There is a tutorial for this tool.

- **NAEP State Profiles** (http://nces.ed.gov/nationsreportcard/states/): The State Profiles present key data about each state's student and school population and its NAEP testing history and results. The profiles also contain links to other sources of information on the NAEP web site, including the most recent state report cards for all available subjects.

- **NAEP Data Explorer (NDE)** (http://nces.ed.gov/nationsreportcard/naepdata/): With the NDE tool users can create statistical tables, charts, and maps to help find answers. They can explore the results of decades of assessment of students' academic performance, as well as information about factors that may be related to their learning.

Program for International Student Assessment (PISA)

- **IDE for PISA—**See IDE, above.

Progress in International Reading Literacy Study (PIRLS)

- **ECB for PIRLS**—Public-use and Restricted-use

- **IDE for PIRLS—**See IDE, above.

Trends in International Mathematics and Science Study (TIMSS)

- **ECB for TIMSS**—Public-use and Restricted-use

- **IDE for TIMSS—**See IDE, above.

- **TIMSS Videotape Classroom Study CD-ROM**: Actual footage of 8th-grade mathematics classes lets viewers see firsthand an abbreviated geometry and algebra lesson in each of three countries: Germany, Japan, and the United States.

HOUSEHOLD SURVEYS

National Household Education Surveys (NHES) Program

- **DAS for NHES—**See DAS, above.

- **ECB for NHES**—Public-use and Restricted-use

SMALL SPECIAL-PURPOSE NCES SURVEYS

High School Transcript Study

- **ECB for 1998 High School Transcript (HST) Study**—Restricted-use

Appendix D: NCES Survey Web Site Addresses

Every effort has been made to verify the accuracy of all URLs listed in this Handbook at the time of publication. If a URL is no longer working, try using the root directory to search for a page that may have moved. For example, if the link to http://nces.ed.gov/surveys/libraries/academic.asp is not working, try http://nces.ed.gov/ and search the NCES Site Index for Academic Libraries.

Survey	Web site
Academic Libraries Survey (ALS)	http://nces.ed.gov/surveys/libraries/academic.asp
Adult Literacy and Lifeskills (ALL) Including the International Adult Literacy Survey (IALS)	http://nces.ed.gov/surveys/all/
Baccalaureate and Beyond (B&B) Longitudinal Study	http://nces.ed.gov/surveys/b%26b/
Beginning Postsecondary Students (BPS) Longitudinal Study	http://nces.ed.gov/surveys/bps/
Common Core of Data (CCD)	http://nces.ed.gov/ccd/
Current Population Survey (CPS), October Supplement	http://nces.ed.gov/surveys/cps/
Early Childhood Longitudinal Study (ECLS)	http://nces.ed.gov/ecls/
Education Longitudinal Study of 2002 (ELS:2002)	http://nces.ed.gov/surveys/els2002/
Fast Response Survey System (FRSS)	http://nces.ed.gov/surveys/frss/
High School and Beyond (HS&B) Longitudinal Study	http://nces.ed.gov/surveys/hsb/
High School Transcript (HST) Studies	http://nces.ed.gov/surveys/hst/
Integrated Postsecondary Education Data System (IPEDS)	http://nces.ed.gov/ipeds/

Survey	Web site
National Assessment of Educational Progress (NAEP)	http://nces.ed.gov/nationsreportcard/
National Assessments of Adult Literacy (NAAL), including the 1992 National Adult Literacy Survey (NALS).	http://nces.ed.gov/naal/
National Education Longitudinal Study of 1988 (NELS:88)	http://nces.ed.gov/surveys/nels88/
National Household Education Surveys (NHES) Program	http://nces.ed.gov/nhes/
National Longitudinal Study of the High School Class of 1972 (NLS-72)	http://nces.ed.gov/surveys/nls72/
National Postsecondary Student Aid Study (NPSAS)	http://nces.ed.gov/surveys/npsas/
National Study of Postsecondary Faculty (NSOPF)	http://nces.ed.gov/surveys/nsopf/
Progress in International Reading Literacy Study (PIRLS)	http://nces.ed.gov/surveys/pirls/
Postsecondary Education Quick Information System (PEQIS)	http://nces.ed.gov/surveys/peqis/
Private School Universe Survey (PSS)	http://nces.ed.gov/surveys/pss/
Program for International Student Assessment (PISA)	http://nces.ed.gov/surveys/pisa/
School Crime Supplement (SCS)	See http://bjs.ojp.usdoj.gov/
School Survey on Crime and Safety (SSOCS)	http://nces.ed.gov/surveys/ssocs/
Schools and Staffing Survey (SASS)	http://nces.ed.gov/surveys/sass/
SASS School Library Media Survey	http://nces.ed.gov/surveys/libraries/school.asp
Survey of Earned Doctorates (SED)	http://www.nsf.gov/statistics/srvydoctorates/
Teacher Follow-up Survey (TFS)	See http://nces.ed.gov/surveys/sass/
Trends in International Mathematics and Science Study (TIMSS)	http://nces.ed.gov/timss/

Appendix E: Index

The Index entries include survey component names and the words defined in the Key Concepts sections. Survey component names are italicized, and the page number where the component description appears in the Overview section is also italicized. Words defined in the Key Concepts sections are identified by an asterisk, and an asterisk follows the page number where the definition appears.

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

T

U

V

W