National Center for Education Statistics

Education Longitudinal Study of 2002: Base Year Data File User's Manual

User's Manual

U.S. Department of Education
Institute of Education Sciences
NCES 2004–405
(Revised)

This page is intentionally blank.

National Center for Education Statistics

Education Longitudinal Study of 2002: Base Year Data File User's Manual

User's Manual

U.S. Department of Education
Institute of Education Sciences
NCES 2004–405
(Revised)

February 2004

Steven J. Ingels
Daniel J. Pratt
James E. Rogers
Peter H. Siegel
Ellen S. Stutts
RTI International

Jeffrey A. Owings
Project Officer
National Center for Education Statistics

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Robert Lerner
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.

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:

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

February 2004

The NCES World Wide Web Home Page address is *http://nces.ed.gov*
The NCES World Wide Web Electronic Catalog is: *http://nces.ed.gov/pubsearch*

Suggested Citation

U.S. Department of Education, National Center for Education Statistics. *Education Longitudinal Study of 2002: Base Year Data File User's Manual*, NCES 2004–405, by Steven J. Ingels, Daniel J. Pratt, James E. Rogers, Peter H. Siegel, and Ellen S. Stutts. Project Officer: Jeffrey A. Owings. Washington, DC: 2004.

For ordering information on this report, write:
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

Or call toll free 1–877–4ED–Pubs

Content Contact:
Jeffrey A. Owings
(202) 502–7423
Jeffrey.Owings@ed.gov

Foreword

This manual has been produced to familiarize data users with the procedures followed for data collection and processing for the base year of the Education Longitudinal Study of 2002 (ELS:2002). It also provides the necessary documentation for use of the public-use data files, as they appear on the ELS:2002 base year Electronic Codebook (ECB).

Analysts do not need to be sophisticated statisticians or computer programmers to use the ELS:2002 ECB. Most social scientists and policy analysts should find the data set organized and equipped in a manner that facilitates straightforward production of statistical summaries and analyses. This manual provides extensive documentation of the content of the data files and how to access and manipulate them.

Chapter 1 serves as an introduction to ELS:2002. It includes an overview and history of the National Center for Education Statistics (NCES) program of longitudinal high school cohorts, summarizes the ELS:2002 objectives, and supplies an overview of the base year and longitudinal study design.

Chapter 2 describes the data collection instruments, including both the development and content of the student, parent, school administrator, teacher, and library media center questionnaires, as well as the student assessments in reading and mathematics, and the facilities checklist.

The sample design and weighting procedures used in the base year study are documented in chapter 3, as are weights, imputation, and the calculation of design effects.

Data collection schedules, training, procedures, and results are presented in chapter 4. Chapter 5 describes data preparation and processing, including the receipt control system, optical scanning, machine editing, and data file preparation.

Chapter 6 describes the contents of the data files, including the data structure and analysis populations.

The appendices include, among other topics, an introduction to the public-use ECB (appendix A), a glossary of special terms used in the ELS:2002 documentation (appendix E), and a crosswalk to the National Education Longitudinal Study of 1988 (NELS:88) and the High School and Beyond (HS&B) longitudinal study sophomore questionnaires (appendix H). Three additional appendices are available online only as PDF files: the ELS:2002 questionnaires (appendix B); a hardcopy codebook with response frequencies, percents, and weighted percents (appendix G); and tables of bias estimates for high nonresponse variables, based on selected key school and student characteristics (appendix I).

This page is intentionally blank.

Acknowledgments

Daniel J. Pratt of RTI served as the ELS:2002 base year project director. Steven J. Ingels of RTI was principal investigator. Jeffrey A. Owings served as the NCES project officer. Key RTI task leaders were Ellen Stutts (data collection task leader, associate project director), James Rogers (data processing), and Peter Siegel (sampling and statistics). Other RTI staff who played major roles in the ELS:2002 base year study were Christopher Alexander, Kimberly Ault, Laura J. Burns, James Chromy, Priscilla Collinson, Elizabeth Copello, George Dunteman, Brian Evans, Deborah Herget, Sheila Hill, Mani Medarametla, Andreina Perez-Greene, Donna Jewell, Ruby Johnson, Amy Rees Sommer, Milorad Stojanovic, Brian Sutton, and Donghui Wang. Assessment development, scaling, and equating were conducted by Judith M. Pollack, Donald A. Rock, and Michael Weiss, under a subcontract with Educational Testing Service (ETS). ETS staff contributed assessment documentation to this manual. Phillip Kaufman directed a further subcontract, at MPR Associates; MPR assisted in preparing the base year statistical analysis report, *Profile of the American High School Sophomore in 2002*. Martin R. Frankel of Abt Associates served the project as a statistical advisor, and Ronald Hambleton of the University of Massachusetts provided psychometric consultation.

The authors of this user's manual would like to thank the many individuals who assisted in the planning of ELS:2002. We are particularly indebted to the ELS:2002 Technical Review Panel, whose members reviewed plans for the study, helped refine them, and provided important suggestions to help guide development of the instrumentation. The following individuals serve as members of the ELS:2002 Technical Review Panel: Marco Clark, Richard Duran, Jeremy Finn, Thomas B. Hoffer, Thomas Kane, Sally Kilgore, Richard Lawrence, Samuel R. Lucas, Aaron Pallas, and Andy Rogers.

Special thanks are in order to Jeffrey A. Owings, Associate Commissioner for Elementary/Secondary & Libraries Studies at NCES, in whose division ELS:2002 is housed. Other NCES staff who have provided help, support, and assistance include Ghedam Bairu, Kathy Chandler, Bill Fowler, Lisa Hudson, Steve Kaufman, Mariann Lemke, Andrew G. Malizio, Edith McArthur, and Marilyn M. Seastrom, as did former NCES staff Samuel Peng and Ilona Berkovits. Staff in other offices of the U.S. Department of Education who have contributed to the study include Clifford Adelman, Sharon Belli, Jeffery Rodamar, and Marsha Silverberg.

Special thanks are also due to Leslie A. Scott of the Education Statistics Services Institute, who contributed to various aspects of the instrumentation, design, and methodology of the base year study, and to Denise M. Davis, formerly of the National Commission on Libraries and Information Science, who helped in the design of the library media center component. We wish to acknowledge the assistance of Bruce Daniel of Pinkerton Computer Consultants, who generated the item nonresponse bias tables in appendix I.

Our thanks also to the reviewers of this user's manual, who provided many helpful comments. The principal reviewers were Mike Planty and Gerard Rainville of the Education Statistics Services Institute; Tai Phan, Ralph Lee, Jim Griffith, Tracy Hunt-White, Carl Schmitt, Patrick Rooney, and Karen O'Conor of the National Center for Education Statistics; Martin R.

Frankel of Abt Associates; and Jeremy D. Finn, University of Buffalo, State University of New York.

Many others, far too many to name individually, contributed to the ELS:2002 base year study, including survey administrators and other contractor and subcontractor staff, and of course the literally thousands of students, parents, and school personnel who generously gave of their time to provide the data that the study reports. We extend our thanks to all.

Lastly, we want to acknowledge the help of several other individuals at RTI in the preparation of this document: Sallie Fiore, Jill Snider, Valerie Stockett, and Robin Whaley, who edited this manual, Lynne Hawley, Lil Clark, Joyce Clay-Brooks, and Sharon Powell, who implemented the final formatting, and Sonja Douglas, Janet Austin, and Sharon Davis, who provided graphics support.

Table of Contents

Foreword iii

Acknowledgments v

List of Tables xi

List of Figures xiii

Chapter 1 Introduction 1

1.1 Overview of the Data File User's Manual 1
1.2 Historical Background 1
1.2.1 NCES Education High School Longitudinal Studies Program 1
1.2.2 National Longitudinal Study of the High School Class of 1972 (NLS-72) 2
1.2.3 High School and Beyond (HS&B) 3
1.2.4 National Education Longitudinal Study of 1988 (NELS:88) 4
1.2.4.1 Base Year 5
1.2.4.2 First Follow-up 5
1.2.4.3 NELS:88 Second Follow-up 6
1.2.4.4 NELS:88 Third Follow-up 6
1.2.4.5 NELS:88 Fourth Follow-up 6
1.3 Education Longitudinal Study of 2002 (ELS:2002) 7
1.3.1 ELS:2002 Study Objectives 7
1.3.2 ELS:2002 Research and Policy Issues 9
1.3.2.1 Cross-sectional Profiles 11
1.3.2.2 Longitudinal Analysis 11
1.3.2.3 Intercohort Comparisons 11
1.3.2.4 International Comparisons 11
1.3.3 Overview of the Base Year Study Design 12

Chapter 2 Instrumentation 13

2.1 Introduction 13
2.1.1 Instrument Development Process and Procedures 13
2.1.2 Instrument Development Goals and Constraints 14
2.2 Student Questionnaire and Assessments 16
2.2.1 Student Questionnaire 16
2.2.2 Reading and Mathematics Assessments 18
2.2.2.1 Test Design and Format 18
2.2.2.2 Scoring Procedures 19
2.2.2.3 Score Descriptions and Summary Statistics 19
2.2.2.4 Links to Other Surveys 21
2.3 Parent Questionnaire 27
2.4 Teacher Questionnaire 29
2.4.1 Teacher Evaluations of ELS:2002 Students 29
2.4.2 Teacher Background 29
2.5 School Administrator Questionnaire 29
2.5.1 School Characteristics 30

2.5.2 Student Characteristics 30
2.5.3 Teaching Staff Characteristics 30
2.5.4 School Policies and Programs 30
2.5.5 Technology 31
2.5.6 School Governance and Climate 31
2.6 Library Media Center Questionnaire 31
2.7 School Facilities Checklist 32

Chapter 3 Sample Design, Weighting, Design Effects, and Data Quality 33

3.1 Introduction 33
3.1.1 Sample Design 33
3.1.2 Schools 33
3.1.3 Students 34
3.1.4 Parents 35
3.1.5 Teachers 35
3.1.6 Nonresponse Bias Analysis 36
3.1.7 Imputation 36
3.1.8 Weighting 36
3.1.9 Standard Errors and Design Effects 37
3.1.10 Disclosure Analysis 37
3.2 Base Year Sample Design 37
3.2.1 Sampling Frame 37
3.2.2 Stratification and School Selection 40
3.2.3 Selection of Students 44
3.2.3.1 Sample Sizes 44
3.2.3.2 Specifications for Enrollment Lists 46
3.2.3.3 Quality Assurance Checks 47
3.2.3.4 Student Sampling from Lists 48
3.2.3.5 Sample Updating 49
3.2.4 Student Eligibility and Exclusion 52
3.2.4.1 Schools Given Clear Criteria for Including/Excluding Students 52
3.2.4.2 Accommodations Offered to Increase Participation 53
3.2.4.3 Questionnaire Eligibility Status to Be Reassessed in the First Follow-up 54
3.2.4.4 Enrollment Status, Records, and Contextual Data Gathered for Students Unable to Be Surveyed or Validly Assessed 54
3.2.5 Selection of Contextual Samples 55
3.2.6 Bias Analysis for School and Student Unit Nonresponse 56
3.2.7 Questionnaire Item Nonresponse Bias Analysis 70
3.2.7.1 High Nonresponse Questionnaire Variables: Student-Level Items 71
3.2.7.2 High Nonresponse Questionnaire Variables: School-Level Items 73
3.2.7.3 Estimating Bias 74
3.2.7.4 Summarization of Bias: Magnitude of Bias; Characteristics Associated with Bias 75
3.2.8 Other Data Quality Issues: Reports of Hours Spent in Viewing Television and Playing Video Games. 79
3.3 Imputation 80
3.4 Calculation of Weights and Results of Weighting 82
3.4.1 School Weights 82
3.4.2 Student Weights 90
3.4.3 Quality Control 99
3.5 Standard Errors and Design Effects 100
3.5.1 Standard Errors 100

3.5.2 Design Effects 101
3.6 Disclosure Risk Analysis and Protections 108

Chapter 4 Data Collection Methodology and Results 109

4.1 Data Collection Overview 109
4.2 Study Endorsements, School Recruitment, and Pre-data Collection Activities 110
4.2.1 Study Endorsements 110
4.2.2 School Recruitment 111
4.2.3 Pre-data Collection Activities 113
4.3 Data Collection Procedures, Student Questionnaires, and Tests 114
4.4 Data Collection Procedures, Teacher Survey 116
4.5 Data Collection Procedures, School Administrator Survey 117
4.6 Data Collection Procedures, Library Media Center Survey 118
4.7 Data Collection Procedures, Facilities Checklist 118
4.8 Data Collection Procedures, Parent Survey 118
4.9 Data Collection Results 119

Chapter 5 Data Preparation and Processing 123

5.1 Overview of Systems Design, Development, and Testing 123
5.2 Data Receipt 124
5.3 Coding for Hardcopy Instruments 125
5.4 Data Capture for Optically Scanned Instruments 125
5.5 Data Cleaning and Editing 126
5.6 Data Capture and Editing for CATI 127
5.7 Data Processing and File Preparation 127

Chapter 6 Data File Contents 129

6.1 Data Structure 129
6.2 Base Year Analysis Populations 129
6.3 Weights and Flags 130
6.4 Composite and Classification Variables 130
6.5 Naming Conventions 131
6.6 Guide to the Hardcopy Codebooks 131

References 133

Appendix A Introduction to the Electronic Codebook A-1

A.1 Obtaining the ELS:2002 ECB A-3
A.2 Features and Content of the ELS:2002 ECB A-3
A.2.1 Student Megafile A-4
A.2.2 School Megafile A-5
A.3 Installing the ECB A-6
A.3.1 Hardware/Software Requirements A-6
A.3.2 Installation Procedures A-6
A.4 Using the ECB A-6
A.4.1 Understanding the File Structure and Capacity A-6
A.4.2 Examining the Frequencies Available for Each Variable on the ECB A-6
A.4.3 Creating a Taglist, Extracting Data, and Generating Program Code A-8
A.4.4 Variance Estimation A-9
A.5 Additional Sources of Information (NCES Reports, Bibliographic Resources) A-9
A.6 Reference A-9

Appendix B Base Year Questionnaires B-1

Appendix C Student Questionnaire Critical Items C-1

Appendix D Public-Use Masked/Suppressed Variables Available on Restricted Files for Licensed Users D-1

Appendix E Glossary of Terms E-1

Appendix F Documentation for Imputed Variables F-1

Appendix G Base Year Codebooks G-1

Appendix H Cross-Cohort Comparisons H-1

H.1 Cross-Cohort Comparison Crosswalks H-3
H.2 Reference H-34

Appendix I Item Nonresponse Bias Analysis Tables I-1

Appendix J Details of School and Student Sampling J-1

J.1 School Sampling J-3
J.2 Student Sampling J-4
J.3 References J-5

Appendix K Standard Errors and Design Effects K-1

List of Tables

Table 1. Number of items in each ELS:2002 test form for assessing achievement in mathematics and reading: 2002 19

Table 2. Item Response Theory (IRT)-estimated number right scores from ELS:2002 mathematics and reading assessments: 2002 20

Table 3. Standardized scores (T-scores) from ELS:2002 mathematics and reading assessments: 2002 21

Table 4. Quartile scores from ELS:2002 mathematics and reading assessments: 2002 21

Table 5. ELS:2002 Item Response Theory (IRT) NELS-equated estimated number right scores: 2002 22

Table 6. Reading and mathematics probability of NELS-equated proficiency scores: 2002 23

Table 7. ELS:2002 and Program for International Student Assessment:2000 (PISA:2000) samples: 2002 24

Table 8. ELS:2002 and Program for International Student Assessment:2000 (PISA:2000) equating sample: 2002 25

Table 9. Comparative statistics for full-sample Program for International Student Assessment:2000 (PISA:2000) and ELS:2002 base year: 2002 25

Table 10. Linking methods for implementing Program for International Student Assessment:2000 (PISA:2000) reading scales in ELS:2002: 2002 26

Table 11. School sampling, eligibility, and participation, by sampling stratum: 2002 42

Table 12. School overlap between the ELS:2002 and the 2002 NAEP: 2002 43

Table 13. Domain sample size requirements: 2002 45

Table 14. Projected sample sizes, by race/ethnicity domains: 2002 46

Table 15. Types of student lists provided, by schools: 2002 47

Table 16. Types of problems encountered with student lists: 2002 48

Table 17. Expected and achieved student samples, by student stratum: 2002 50

Table 18. Counts of students excluded and students accommodated: 2002 55

Table 19. Sample teachers, by subject taught, school type, and school urbanicity: 2002 56

Table 20. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for schools: 2002 60

Table 21. Nonresponse bias before and after nonresponse adjustment for selected continuous variables for schools: 2002 63

Table 22. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for students: 2002 64

Table 23. Nonresponse bias before and after nonresponse adjustment for selected continuous variables for students: 2002 67

Table 24. Student-level high nonresponse questionnaire variables, by weighted response rate: 2002 .. 71

Table 25. School-level high nonresponse questionnaire variables, by weighted response rate: 2002 ... 73

Table 26. ELS:2002 student file, 78 high nonresponse variable summary, by 40 characteristics: mean, median, and standard deviation of bias estimates: 2002 ... 75

Table 27. Frequency distribution of unsigned bias estimates, 78 high nonresponse student variables, by 40 characteristics: 2002 ... 75

Table 28. Mean, median, and standard deviation for bias estimates for each of 40 characteristics, across 78 high nonresponse student file variables: 2002 ... 76

Table 29. ELS:2002 school file, 41 high nonresponse variable summary, by 10 characteristics: mean, median, and standard deviation of bias estimates: 2002 ... 77

Table 30. Frequency distribution of unsigned bias estimates, 41 high nonresponse school variables, by 10 characteristics: 2002 ... 78

Table 31. Mean, median, and standard deviation for bias estimates for each of 10 characteristics across 41 high nonresponse school file variables: 2002 ... 78

Table 32. ELS:2002 imputation variables: 2002 ... 81

Table 33. Average weight adjustment factors used to adjust school weights for nonresponse: 2002 ... 85

Table 34. Average weight adjustment factors for poststratifying to control totals: 2002 ... 88

Table 35. Statistical properties of school weight: 2002 ... 90

Table 36. Average weight adjustment factors used to adjust student weights for parent refusal: 2002 ... 92

Table 37. Average weight adjustment factors used to adjust student weights for other nonresponse: 2002 ... 96

Table 38. Statistical properties of student weights: 2002 ... 99

Table 39. Mean design effects (DEFFs) and root design effects (DEFTs) for school and library questionnaire data: 2002 ... 103

Table 40. Mean design effects (DEFFs) and root design effects (DEFTs) for student questionnaire data: 2002 ... 103

Table 41. Mean design effects (DEFFs) and root design effects (DEFTs) for parent questionnaire data: 2002 ... 104

Table 42. Summary of ELS:2002 base year completion and coverage rates: 2002 ... 109

Table 43. ELS:2002 base year school sample selections and realization: 2002 ... 120

Table 44. ELS:2002 base year completion/coverage rates, by component at the student level: 2002 ... 121

Table 45. ELS:2002 base year school-level participation rates and completion rates for school surveys at the school level: 2002 ... 122

List of Figures

Figure 1. Longitudinal design for the NCES high school cohorts: 2002 3
Figure 2. Before versus after nonresponse adjustment—school-level relative bias: 2002 68
Figure 3. Before versus after nonresponse adjustment—Student-level relative bias: 2002..... 69
Figure 4. Minimum bias ratio by Type I error rate: 2002 69
Figure 5. HS&B, NELS:88, and ELS:2002 mean design effects and root design effects: 2002 105
Figure 6. Completion and coverage rates for ELS:2002 base year: 2002..... 110
Figure 7. Survey administrator training agenda: 2002..... 114

This page is intentionally blank.

Chapter 1
Introduction

1.1 Overview of the Data File User's Manual

This manual provides guidance and documentation for users of the public release data for the base year of the Education Longitudinal Study of 2002 (ELS:2002). ELS:2002 is sponsored by the National Center for Education Statistics (NCES) of the Institute of Education Sciences, U.S. Department of Education. The base year study was conducted through a contract to RTI International (RTI)[1], a university-affiliated, nonprofit research organization in North Carolina, in collaboration with its subcontractors, the Educational Testing Service of Princeton, New Jersey, and MPR Associates of Berkeley, California. This manual contains information about the purposes of the study, the data collection instruments, the sample design, and data collection and data processing procedures. The manual provides guidance for understanding and using all components of the base year study—student questionnaire and test data; data from parents; data from teachers, school administrators, librarians, and media center specialists; and observational data gathered in the school facilities checklist.

The ELS:2002 base year data set has been produced in both public-use and restricted-use versions (see appendix D for a summary of differences between the public and restricted Electronic Cookbooks). The released data files reflect alteration or suppression of some of the original data. Such edits were imposed to minimize the risk of disclosing the identity of responding schools and the individuals within them. While the primary focus of this manual is the public-release version of the data as issued in Electronic Codebook (ECB) format, much of the information supplied is also applicable to the restricted-use ECB.

Chapter 1 addresses three main topics. First, it supplies an overview of the NCES education longitudinal studies program, thus situating ELS:2002 in the context of the earlier NCES high school cohorts studied in the 1970s, 1980s, and 1990s. Second, it introduces ELS:2002 by delineating its principal objectives. Third, it provides an overview of the base year study design. In subsequent chapters, these additional topics are addressed: instrumentation (chapter 2), sample design and weighting (chapter 3), data collection methods and results (chapter 4), data preparation and processing (chapter 5), and data file contents (chapter 6). Appendices provide additional information, including a brief introduction to the base year ECB.

1.2 Historical Background

1.2.1 NCES Education High School Longitudinal Studies Program

In response to its mandate to "collect and disseminate statistics and other data related to education in the United States" and the need for policy-relevant, nationally representative longitudinal samples of elementary and secondary students, NCES instituted the National Education Longitudinal Studies program. The aim of this continuing program is to study the educational, vocational, and personal development of students at various stages in their

[1] RTI International is a trade name of Research Triangle Institute.

educational careers, and the personal, familial, social, institutional, and cultural factors that may affect that development.

NCES (and ELS:2002) are authorized by section 406(b) of the General Education Provision Act (20 U.S.C. 1221e) as amended by the Education Sciences Reform Act of 2002. The Education Sciences Reform Act of 2002 replaced the former Office of Educational Research and Improvement (OERI) with the Institute of Education Sciences (IES), in which NCES is now housed.

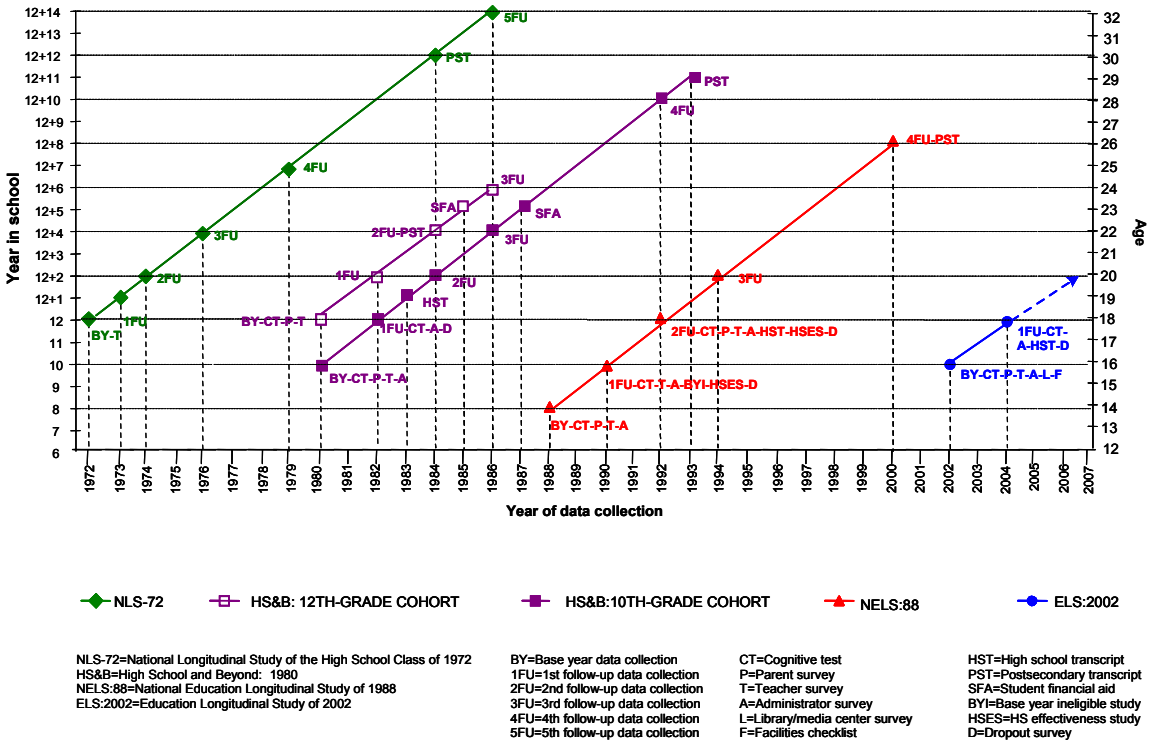
The high school longitudinal studies program consists of three completed studies: the National Longitudinal Study of the High School Class of 1972 (NLS-72), the High School and Beyond (HS&B) longitudinal study of 1980, and the National Education Longitudinal Study of 1988 (NELS:88). In addition, base year data for ELS:2002, the fourth longitudinal study in the series, are now available. Taken together, these studies describe (or will describe) the educational experiences of students from four decades—the 1970s, 1980s, 1990s, and 2000s—and also provide bases for further understanding of the correlates of educational success in the United States. Figure 1 includes a temporal presentation of these four longitudinal education studies and highlights their component and comparison points. Figure 1 does not identify all future follow-up points for ELS:2002; final decisions have yet to be made concerning them. However, the general expectation is that ELS:2002 sophomores will be followed until about age 30.

1.2.2 National Longitudinal Study of the High School Class of 1972 (NLS-72)

The Education Longitudinal Studies program began over 30 years ago with the implementation of NLS-72.[2] NLS-72 was designed to provide longitudinal data for educational policymakers and researchers who link educational experiences in high school with important downstream outcomes such as labor market experiences and postsecondary education enrollment and attainment. With a national probability sample of 19,001 high school seniors from 1,061 public and religious and other private schools, the NLS-72 sample was representative of approximately 3 million high school seniors enrolled in 17,000 U.S. high schools during the spring of the 1971–72 school year. Each member of this cohort was asked to complete a student questionnaire and a cognitive test battery. In addition, administrators at the sample members' schools were asked to supply information about the schools' programs, resources, and grading systems, as well as survey data on each student. No parent survey was conducted. However, postsecondary education transcripts were collected from the institutions attended by students. Five follow-up surveys were completed with this student cohort, with the final data collection taking place in 1986, when the sample members were 14 years removed from high school and approximately 32 years old.

[2] For reports on the NLS-72 project, see Riccobono, Henderson, Burkheimer, Place, and Levinsohn (1981) and Tourangeau, Sebring, Campbell, Glusberg, Spencer, and Singleton (1987). While recent NCES reports and user documentation may be found on the NCES web site (*http://nces.ed.gov*), older documentation (e.g., from the 1980s) is typically not available there. NLS-72 and older HS&B manuals may be downloaded from the International Archive of Education Data (IAED) at the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan (*http://www.icpsr.umich.edu*). Materials may also be obtained in microfiche or photocopy format from ERIC (*http://www.askeric.org/*).

Figure 1. Longitudinal design for the NCES high school cohorts: 2002



A wide variety of data were collected in the NLS-72 surveys. For example, in addition to background information about the student and his or her family, the base year and follow-up surveys collected data on each respondent's educational activities (e.g., schools attended, grades received, and degree of satisfaction with educational institutions). Participants were also asked about their work experiences, periods of unemployment, job satisfaction, military service, marital status, and children. Attitudinal information on self-concept, goals, community involvement, and personal evaluations of educational activities were also included in the study.

1.2.3 High School and Beyond (HS&B)

The second in the series of NCES longitudinal studies was launched in 1980. HS&B included one cohort of high school seniors comparable to the NLS-72 sample; however, the study also extended the age span and analytical range of NCES longitudinal studies by surveying a sample of high school sophomores. Base year data collection took place in the spring term of the 1979–80 academic year with a two-stage probability sample. More than 1,000 schools served as the first-stage units, and 58,000 students within these schools were the second-stage units. Both cohorts of HS&B participants were resurveyed in 1982, 1984, and 1986; the sophomore group also was surveyed in 1992.[3] In addition, to better understand the school and home contexts for the sample members, data were collected from teachers (a teacher comment form in the base year asked for teacher perceptions of HS&B sample members), principals, and a

[3] For a summation of the HS&B sophomore cohort study, see Zahs, Pedlow, Morrissey, Marnell, and Nichols (1995). For more information on HS&B in the high school years, with a focus on the sophomore cohort, see Jones, Clarke, Mooney, McWilliams, Crawford, Stephenson, and Tourangeau (1983). For further information on HS&B, see the NCES web site: *http://nces.ed.gov/surveys/hsb/*.

subsample of parents. High school transcripts were collected for a subsample of sophomore cohort members. As in NLS-72, postsecondary transcripts were collected for both HS&B cohorts; however, the sophomore cohort transcripts cover a much longer time span (to 1993).

With the study design expanded to include a sophomore cohort, HS&B provided critical data on the relationships between early high school experiences and students' subsequent educational experiences in high school. For the first time, national data were available that showed students' academic growth over time and how family, community, school, and classroom factors promoted or inhibited student learning. Researchers were able to use data from the extensive battery of achievement tests within the longitudinal study to assess growth in knowledge and cognitive skills over time. Moreover, data were then available to analyze the school experiences of students who later dropped out of high school, and eventually, to investigate their later educational and occupational outcomes. These data became a rich resource for policymakers and researchers over the next decade and provided an empirical base to inform the debates of the educational reform movement that began in the early 1980s.[4]

1.2.4 National Education Longitudinal Study of 1988 (NELS:88)

Much as NLS-72 captured a high school cohort of the 1970s and HS&B captured high school cohorts of the 1980s, NELS:88 was designed to study high school students of the 1990s—but with a premeasure of their achievement and status, prior to their entry into high school. NELS:88 represents an integrated system of data that tracked students from junior high or middle school through secondary and postsecondary education, labor market experiences, and marriage and family formation. Because ELS:2002 repeats so many of its innovations and design features, it will be useful to provide a detailed round-by-round picture of NELS:88.

Data collection for NELS:88 was initiated with the eighth-grade class of 1988 in the spring term of the 1987–88 school year. Along with a student survey, NELS:88 included surveys of parents (base year and second follow-up), teachers (base year, first, and second follow-ups), and school administrators (base year, first, and second follow-ups). The sample was also surveyed after scheduled high school graduation, in 1994 and 2000.[5]

[4] For a summary of reforms instituted between the time the HS&B cohort was in high school and the NELS:88 cohort was in middle/junior high and high school, see Rasinski, Ingels, Rock, and Pollack (1993) or Barton and Coley (1990). For a summary of state education reforms instituted during the earlier school years of the ELS:2002 cohort, see Hurst, Tan, Meek, and Sellers (2003).

[5] The entire compass of NELS:88, from its baseline through its final follow-up in 2000, is described in Curtin, Ingels, Wu, and Heuer (2002). More detailed information about the sophomore surveys of NELS:88 can be found in Ingels, Scott, Rock, Pollack, and Rasinski (1994). Outcomes for the eighth-grade cohort in 2000 are reported in Ingels, Curtin, Kaufman, Alt, and Chen (2002). The most extensive documentation of the NELS:88 assessment battery is found in Rock and Pollack (1995). The quality of NELS:88 data in the in-school rounds is examined in McLaughlin and Cohen (1997). The sample design is documented in Spencer, Frankel, Ingels, Rasinski, and Tourangeau (1990). Eligibility and exclusion issues are addressed in Ingels (1996). NCES keeps an updated version of the NELS:88 bibliography on its web site. The bibliography encompasses both project documentation and research articles, monographs, dissertations, and paper presentations employing NELS:88 data (see *http://nces.ed.gov/surveys/nels88/Bibliography.asp*).

1.2.4.1 Base Year

The NELS:88 base year (1988) successfully surveyed 24,599 students, out of some 26,432 selected eighth graders, across 1,052 public, Catholic, and other private schools. In addition to filling out a questionnaire, students also completed assessments in four subjects (the NELS:88 achievement battery included tests in reading, mathematics, science, and social studies). The base year also surveyed one parent, two teachers, and the principal of each selected student. The base year research instruments collected information about home, school, and individual factors that could serve as predictors for later outcomes (such as, viewed in terms of positive outcomes, graduating from high school, making a smooth transition into the work force, or completing postsecondary education). Information collected in the base year included family income, parental education and occupation; parental aspirations for their eighth grader; the eighth grader's educational and occupational aspirations and plans, school experiences, extracurricular activities, jobs and chores, television viewing, and reading; teacher perceptions of the eighth grader's classroom performance and personal characteristics; curricular and instructional information about the classes in which teachers taught the eighth grader; the teacher's own background and activities; and the principal's reports on the educational setting and environment of the school.

1.2.4.2 First Follow-up

A first follow-up took place in 1990. At that time, student cohort members, their teachers, and their principals were resurveyed. The first follow-up presented three major new analytic opportunities: (1) longitudinal analysis of gains in tested achievement and the correlates of achievement gains, (2) identification of high school dropouts and investigation of why some students drop out of school and others persist, and (3) cross-cohort comparison (1990 high school sophomores could be compared to sophomores in 1980).

Achievement Gain. One major goal of NELS:88 was to measure students' academic growth over time and to identify the specific school (and nonschool) processes that foster academic achievement. The first follow-up tests were tailored to students' ability as measured in the base year; more difficult test forms were assigned to students with a higher ability estimate. The first follow-up, by retesting the NELS:88 eighth-grade cohort, was able to measure cognitive gains between eighth and 10th grades in mathematics, science, reading, and social studies. In turn, these gains could be related to the data collected on home and school influences on achievement, starting in 1988. Because NELS:88 developed hierarchical criterion-referenced proficiency scores (in reading, science, and mathematics), gain can be looked at in more than just quantitative terms—one can use the proficiency levels to locate the place on the growth continuum where the gain took place (e.g., at a lower or at a higher skill area) and, in turn, better relate gains to specific school processes and curricular sequences.[6]

Determinants and Dynamics of Dropping Out. Another major goal of the first follow-up was to study the educational trajectory of those who drop out of high school and to better understand the factors that help some at-risk students persist in their education. By beginning with the eighth grade, NELS:88 was able to capture the population of early dropouts—those who

[6] Further information about NELS:88 proficiency scores can be found in Rock and Pollack (1995a). For examples of their use in achievement gain analysis, see Rock and Pollack (1995b) and Scott, Rock, Pollack, and Ingels (1995).

left school prior to spring term of 10th grade—as well as (in the second follow-up) later dropouts (who left after spring of 10th grade) as had been studied in HS&B.

Cross-cohort Comparison. A third goal of the 1990 wave was to compare NELS:88 sophomores with the earlier cohort of high school sophomores studied in HS&B. To ensure comparability of the two samples, NELS:88 had to "freshen" the sophomore sample by giving a chance of selection to 1990 sophomores who had not been eighth graders in 1988 (or had not been in the United States). Thus, a nationally representative sophomore grade cohort was included in NELS:88 in the first follow-up (1990). The freshening of the sample provided comparability to earlier cohorts and opportunities for comparing the situation of NELS:88 sophomores with those of HS&B a decade before. Freshening also enabled researchers to conduct both grade-representative cross-sectional and subsequent sophomore cohort longitudinal analyses with the data.

1.2.4.3 NELS:88 Second Follow-up

The second follow-up took place in the spring term of the 1991–92 school year, when most sample members were in their final semester of high school. There were 21,188 student and dropout participants. This follow-up provided a culminating measurement of learning in the course of secondary school and also collected information to facilitate investigation of the transition into the labor force and postsecondary education after high school. As in the first follow-up, the sample was freshened, this time to represent the high school senior class of 1992. Trend comparisons can be made to the high school classes of 1972 and 1980 that were studied in NLS-72 and HS&B. The NELS:88 second follow-up also surveyed students who were identified as dropouts in 1990 and identified and surveyed additional students who had left school since the prior wave. In late 1992 and early 1993, high school transcripts were collected for sample members.

1.2.4.4 NELS:88 Third Follow-up

The third follow-up took place in 1994, when most sample members had completed high school. The primary goals of the 1994 round were (1) to provide data for trend comparisons with NLS-72 and HS&B, (2) to address issues of employment, (3) to address issues of postsecondary access and choice, and (4) to ascertain how many dropouts had returned to school and by what route. There were 14,915 participants.

1.2.4.5 NELS:88 Fourth Follow-up

The fourth follow-up took place in 2000, when most sample members who attended college and technical schools had completed their postsecondary education. The study data address issues of employment, family formation, and postsecondary persistence and attainment. There were 12,144 participants in the questionnaire phase of the study. In fall 2000 and early 2001, postsecondary transcripts were collected, further increasing the analytic potential of the data and the possibility of examining trends over time.

1.3 Education Longitudinal Study of 2002 (ELS:2002)

The base year of ELS:2002 represents the first stage of 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 will be 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 work force.

This section introduces ELS:2002, lists some of the major research and policy issues that the study addresses, and explains the four levels of analysis—cross-sectional, longitudinal, cross-cohort, and international comparison—that can be conducted with ELS:2002 data.

1.3.1 ELS:2002 Study Objectives

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.

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 will be surveyed twice, in 2002 (completed) and in 2004. 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, and their schools (from which data have been collected at three levels: from the principal, the librarian, and from a facilities checklist). Each of these two features—the longitudinal nature of the ELS:2002 design and its multilevel focus—will be explained in greater detail below.

The transition through high school and beyond into postsecondary institutions and the labor market is both complex (youth may follow many different pathways) and prolonged (it takes place over a period of years). The complexity and time frame for this transition make longitudinal approaches especially appropriate. By surveying the same young people over time, it is possible to record the changes taking place in their lives. It is also possible to explain these changes, that is, to understand the ways that earlier achievements, aspirations, and experience predict and influence what happens to the respondents later. In the first year of data collection (the 2002 base year), ELS:2002 measured students' tested achievement in reading and mathematics. ELS:2002 also obtained information from students about their attitudes and experiences. These same students will be tested and surveyed again in 2 years' time to measure changes such as achievement gains in mathematics and changes in enrollment status, such as the situation of students who drop out of school as contrasted to those who persist in their education. Cohort members will be followed for a number of years thereafter so that later outcomes (e.g., their access to and persistence in higher education, or their success in the labor market) can be understood in terms of their earlier aspirations, achievement, and high school situation.

ELS:2002 gathers information at multiple levels. It obtains information not only from students and their school records, but also from students' parents, teachers, and the

administrators (principal and library media center director) of their schools. Data from their teachers, for example, provide information both about the student and the teacher's backgrounds and activities. This multilevel focus supplies researchers with a comprehensive picture of the home, community, and school environments and their influences on the student. This multiple respondent perspective is unified by the fact that, for most purposes, the student is the basic unit of analysis.[7]

Using this multilevel and longitudinal information, 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 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, researchers can use ELS:2002 to measure the effects of their 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 work force (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.

Key elements in the ELS:2002 longitudinal design are summarized by wave below.

Base Year (2002)

- Completed baseline survey of high school sophomores in spring term 2002.
- Completed cognitive tests in reading and mathematics.
- Completed survey of parents, English teachers, and mathematics teachers. Collected school administrator questionnaires.
- Included additional components for this study–a school facilities checklist and a media center (library) questionnaire.
- Established sample sizes of approximately 750 schools and over 17,000 students. Schools are the first-stage unit of selection, with sophomores randomly selected within schools.
- Oversampled Asian and Hispanic students and private schools.

[7] Base year school administrator, library media center, and facilities data can be used to report on the nation's schools with 10th grades in the 2001–02 school year. However, the primary use of the school-level data (and the purpose of parent and teacher surveys) is to provide further contextual information on the student.

- Designed linkages with the Program for International Student Assessment (PISA) and the National Assessment of Educational Progress (NAEP); scored reporting linkages to the prior longitudinal studies.

First Follow-up (2004)

- Note that most sample members are seniors, but some are dropouts or in other grades.
- Administer student questionnaire, dropout questionnaire, assessment in mathematics, and school administrator questionnaire to be administered.
- Return to the same schools, but separately follow transfer students.
- Freshen for a senior cohort.
- High school transcript component in 2004 (coursetaking records for grades 9–12 at minimum).

Second Follow-up (2006)

- Post-high-school follow-ups by computer-assisted telephone interview (CATI).
- Survey 2 years after scheduled high school graduation.

Further Follow-ups

- Determine number of (and dates for) further CATI follow-ups.

1.3.2 ELS:2002 Research and Policy Issues

Apart from helping to describe the status of high school students and their schools, ELS:2002 will provide information to help address a number of key policy and research questions. The study is intended to produce a comprehensive data set for the development and evaluation of educational policy at all government levels. Part of its aim is to inform decision makers, educational practitioners, and parents about the changes in the operation of the educational 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. Such subgroups include:

  - students in public and private high schools;
  - language minority students;
  - students with disabilities;
  - students in urban, suburban, and rural settings;
  - students in different regions of the country;
  - students from upper, middle, and lower socioeconomic status levels;
  - male and female high school students; and
  - students from different racial or ethnic groups.
- 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. These issues include:

- the later educational and labor market activities of high school dropouts;
- the transition of those who do not go directly on to postsecondary education or to the world of work;
- access to and choice of, undergraduate and graduate educational 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 work force;
- social and economic rate of return on education to both the individual and society; and
- adult roles, such as family formation and civic participation.

These research and policy issues can be investigated at several distinct levels of analysis. The overall scope and design of the study provide for the four following analytical levels:

- cross-sectional profiles of the nation's high school sophomores and seniors (as well as dropouts after the spring term of their sophomore year);
- longitudinal analysis (including examination of life course changes);
- intercohort comparisons with American high school students of earlier decades; and
- international comparisons: U.S. 15-year-olds to 15-year-olds in other nations.

1.3.2.1 Cross-sectional Profiles

Cross-sectional data will permit characterization of the nation's high school sophomores in the spring term of the 2001–02 school year. Initial cross-sectional findings from the base year are available in an NCES report *A Profile of the American High School Sophomore in 2002*.[8] Because of sample freshening, the results 2 years later will provide a basis for profiling the nation's high school seniors in the spring term of the 2003–04 school year.

1.3.2.2 Longitudinal Analysis

Longitudinal analysis will become possible when data are available from the 2004 first follow-up. The primary research objectives of ELS:2002 are longitudinal in nature. The study provides the basis for within-cohort comparison by following the same individuals over time to measure achievement growth in mathematics, monitor enrollment status over the high school years, and record such key outcomes as postsecondary entry and attainment, labor market experiences, and family formation. In turn, these outcomes can be related to antecedents identified in earlier rounds, including individual, home, school, and community factors.

1.3.2.3 Intercohort Comparisons

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. A 1980–2002 NCES high school sophomore trend report is currently in preparation. With completion of the first follow-up in 2004, researchers will be able to compare ELS:2002 high school seniors' experience, attitudes, and achievement with that of NELS:88 seniors in 1992, HS&B seniors in 1980, and NLS-72 seniors in 1972. They will also be able to compare ELS:2002 dropouts in 1984 with the high school dropouts studied by HS&B in 1982 and by NELS:88 in 1992. Such cross-cohort comparisons are of particular importance to measuring the nation's goals in achieving equity in educational opportunities and outcomes and in measuring the success of school reform and related initiatives.

Starting with the ELS:2002 first follow-up, trend comparisons can also be made with academic transcript data containing students' high school course histories and sequences, since comparable transcript studies have been conducted, starting with HS&B (1982) and including NELS:88 (1992) and NAEP (1987, 1990, 1994, 1998, and 2000).

1.3.2.4 International Comparisons

A feature of ELS:2002 that expands the study's power beyond that of the predecessor studies is that it will be used to support international comparisons. Items have been included on the ELS:2002 achievement tests from the Program for International Student Assessment (PISA). The Organization for Economic Cooperation and Development's (OECD's) PISA[9] is an internationally standardized assessment, jointly developed by the 32 participating countries

[8] See Ingels, Burns, Chen, Cataldi, and Charleston (2004).
[9] See Lemke, Calsyn, Lippman, Jocelyn, Kastberg, Liu, Williams, Kruger, and Bairu (2001).

(including the United States) and administered to 15-year-olds in groups in their schools. PISA covers three domains: reading literacy, numeracy, and scientific literacy—a subset of the PISA reading literacy and numeracy items have been included on ELS:2002. PISA aims to define each domain not merely in terms of mastery of the school curriculum, but also in terms of important knowledge and skills needed in adult life. Emphasis is placed on the mastery of processes, the understanding of concepts, and the ability to function in various situations within each domain.

1.3.3 Overview of the Base Year Study Design

ELS:2002 was carried out in a national probability sample of 752 public, Catholic, and other private schools in the spring term of the 2001–02 school year. Of 17,591 eligible selected sophomores, 15,362 completed a base year questionnaire, as did 13,488 parents, 7,135 teachers, 743 principals, and 718 librarians.

Seven study components comprise the base year design: assessments of students (achievement tests in mathematics and reading); a survey of students; surveys of parents, teachers, school administrators, and librarians; and a facilities checklist (completed by survey administrators, based on their observations at the school). The student assessments measured achievement in mathematics and reading; the baseline scores can serve as a covariate or control variable for later analyses. Mathematics achievement will be reassessed 2 years hence, so that achievement gain over the last 2 years of high school can be measured and related to school processes and mathematics coursetaking. 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.

One parent of each participating sophomore was asked to respond to a parent survey. The parent questionnaire was designed to gauge parental aspirations for their child, home background and the home education support system, the child's educational history prior to 10th grade, and parental interactions with and opinions about the student's school. For each student enrolled in English or mathematics, a teacher was also selected to participate in a teacher survey. The teacher questionnaire collected the teacher's evaluations of the student and provided information about the teacher's background and activities. The head librarian or media center director at each school was asked to complete a library media center questionnaire, which inquired into the school's library media center facility, its staffing, its technological resources, collection and expenditures, and scheduling and transactions. Finally, the facilities checklist was a brief observational form completed for each school. The form collected information about the condition of school buildings and facilities. Information about coursetaking (covering all years of high school and including the sequence in which courses were taken and grades earned) will be collected at the end of high school through the high school transcript component of the ELS:2002 first follow-up study.

Further details of the instrumentation, sample design, data collection results, and the data files available for analysis are found in the chapters that follow.

Chapter 2
Instrumentation

2.1 Introduction

The data collection instruments for the Education Longitudinal Study of 2002 (ELS:2002) base year 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). The base year questionnaires can be found in the electronic version of this data file user's manual (appendix B) as PDF files on the NCES ELS:2002 web site (*http://nces.ed.gov/surveys/els2002/*).

2.1.1 Instrument Development Process and Procedures

Content specification documents were commissioned for the planned achievement tests in reading and mathematics as well as for the student, parent, teacher, and school administrator survey questionnaires. These documents provided an instrument development framework by identifying the key ELS:2002 research questions, the constructs that had to be considered in answering the research questions, and the variables or data elements that could help to inform each construct. The content specification documents drew heavily on existing item pools (e.g., National Assessment of Educational Progress [NAEP], National Education Longitudinal Study of 1988 [NELS:88], and the Program for International Student Assessment [PISA] for the achievement tests; and NELS:88 for the questionnaires).

In general, the development and review process for each questionnaire consisted of the following steps:

1. *Sharing of Draft Data Elements.* Draft elements of the questionnaires were shared with other government agencies, policy groups, and interested parties.
2. *Technical Review Panel (TRP) Review.* The ELS:2002 TRP, a specially appointed, independent group of substantive, methodological, and technical experts, reviewed the questionnaires.
3. *National Center for Education Statistics (NCES) Review.* The questionnaires underwent interdivisional review at NCES.
4. *Questionnaire Revision.* The survey instruments were revised based on reviewer comments.
5. *Writing of Justification.* A justification was written for components of the instruments.
6. *Office of Management and Budget (OMB) Review*. The federal OMB reviewed the instruments.
7. *Questionnaire Revision.* The questionnaires were revised based on OMB comments.
8. *Field Testing and Revision.* The instruments were field tested and revised based on field test results.

Specific assessment items for the reading and mathematics tests were typically not subject to these reviews, but the larger assessment framework and goals, and the results (as seen in overall item statistics from the field test) were an integral element within the review process and, in particular, the deliberations of the TRP.

The field testing of school enlistment and data collection and processing procedures, questionnaires, and assessments was an especially important step in the development of the full-scale base year study. Field test instruments were evaluated in a number of ways. For the questionnaires, field test analyses included evaluation of item nonresponse, examination of test-retest reliabilities, calculation of scale reliabilities, and examination of correlations between theoretically related measures. For the achievement tests in mathematics and reading, item parameters were estimated for both 10th and 12th grade. Both classical and Item Response Theory (IRT) techniques were employed to determine the most appropriate items for inclusion in the final (base year main study) forms of the two tests. Psychometric analyses included various measures of item difficulty and discrimination, investigation of reliability and factor structure, and analysis of differential item functioning. The field test report is available from NCES.[10]

2.1.2 Instrument Development Goals and Constraints

ELS:2002 is a longitudinal study in which data across various waves of data collection are used in analyses. Since the primary research objectives of ELS:2002 are longitudinal in nature, the first priority was to select the items that would prove most useful in predicting or explaining future outcomes as measured in future survey waves.

The second priority was to obtain needed cross-sectional data, whenever consistent with the longitudinal objectives, particularly data that could be used for intercohort comparison with past studies or linkage to certain current data collection efforts. Wherever possible, all ELS:2002 instruments were designed to provide continuity and consistency with the earlier education longitudinal studies of high school cohorts. Where appropriate, ELS:2002 drew from the National Longitudinal Study of the High School Class of 1972 (NLS-72), the High School and Beyond (HS&B) longitudinal study, and, most particularly, NELS:88. In addition, questionnaire and test items were in some cases drawn from other NCES programs, such as NAEP (especially for the assessments), PISA (for both assessments and psychological scales related to orientation toward learning), the Schools and Staffing Survey (SASS) (particularly but not exclusively for items related to the library media center questionnaire), or the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K) (from which was borrowed the concept of a facilities checklist). Continuity with ELS:2002's historical predecessors and with other NCES survey and assessment programs was pursued to ensure a common standard of measurement that would permit comparisons and increase the usefulness of the ELS:2002 data. Apart from the intercohort or cross-study comparisons that can be sustained through use of the questionnaire and transcript data, ELS:2002 provides equated scores with the testing programs of NAEP, PISA, HS&B, and NELS:88.

While maintaining trend items to support intercohort comparisons was a major aim of instrument development, there was also a need to provide new items to address new areas of policy concern and to reflect recent advances in theory. For example, stress was put on adding

[10] See Burns, Heuer, Ingels, et al. (2003) at the NCES web site (*http://nces.ed.gov/pubsearch/*).

items about educational technology, since computers have become a major factor in learning in recent years. Plans were also made to add psychological scales that reflect recent work in self-efficacy theory and related areas.

Another consideration in the development of the ELS:2002 instruments was the need to obtain factual information from the best source among the various respondent populations. In some cases the decision to go with the best source has also entailed waiting longer to secure the information (e.g., the sophomore questionnaire was not used to collect information on courses taken or grades; academic transcripts are a more reliable source of this information, and they will be collected after students have completed high school.) In most cases, information has been collected from one source only. For example, when it was determined that the best source of information about family composition was the parent, the item was put only on the parent questionnaire. In a few instances, a particular datum was deemed to be of such importance that some redundancy between instruments seemed an acceptable price to pay. For example, while parents are the best source of information about occupation and highest parental educational attainment, the importance of these items was such that they were asked both on the student and parent questionnaires, to increase the number of sample members for whom this information would be available.

Finally, some changes in law regarding question areas that could be asked of students in a school setting under conditions of implied consent had to be taken into account. Specifically, the Protection of Pupil Rights Amendment (PPRA) proscribes collection of information in the following seven areas when minor students are required to participate in a survey, unless prior written parental consent has been obtained:

1. political affiliations or beliefs of the student or the student's parent;
2. mental and psychological problems of the student or the student's family;
3. sexual behavior or attitudes;
4. illegal, antisocial, self-incriminating, or demeaning behavior;
5. critical appraisals of other individuals with whom respondents have close family relationships;
6. legally recognized privileged or analogous relationships, such as those of lawyers, physicians, and ministers; and
7. income.

In addition, when the PPRA was amended in the *No Child Left Behind Act of 2001*, an eighth area was added to the list:

8. religious practices, affiliations, or beliefs of the student or student's parent.

A number of topic areas asked about in prior studies such as HS&B and NELS:88 were therefore dropped from the ELS:2002 student questionnaires, including all items on use of tobacco, alcohol, and drugs, and past and present illegal, sexual, or antisocial behaviors, as well as psychological problems and appraisals of family members. A few additional items retained on the student questionnaire that later raised PPRA concerns were suppressed from the final data

set (this fact accounts for the several gaps in the questionnaire and variable name number sequences for the base year student survey).

Basic elements that are or will be encompassed in the ELS:2002 research instruments can be classified in three broad categories:

- background information (normally collected in the base year only, except for respondents first entering the sample in a later round);
- process information (information about dynamic influences on the student in the home, school, and community environment, as he or she moves through secondary school and beyond into the world of postsecondary education and the adult work force); and
- outcome information (the eventual outcomes of the transition process, including later educational attainment and labor market status). The base year questionnaires are rich in background and process items. The final wave of the study will collect primarily outcome data.

2.2 Student Questionnaire and Assessments

2.2.1 Student Questionnaire

The ELS:2002 student questionnaire was a 45-minute self-administered instrument. Sophomore sample members normally completed the questionnaire in a group administration in the classrooms of their schools. A few students were surveyed outside of school, with a shortened version in a computer-assisted telephone interview (CATI). Assessments in reading and mathematics were given at the same time, in a two-stage process in which the first stage was a routing test. Questionnaire administration is described in chapter 4. The full questionnaire was available only in English, although a shortened Spanish version was also produced.

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.

The locating information section primarily gathered information needed for future follow-up; however, it also elicited data that have been used in the creation of some of the standard classification variables for the study: date of birth, sex, Hispanic ethnicity, race, and Asian or Hispanic subgroup.

By far the longest section of the student questionnaire was the module on school experiences and activities. The principal content strands in this section inquire about school climate, student recognition, school disengagement behaviors (tardiness, classes-cutting, etc.), perception of high school program placement (academic, general, or vocational track), attitudes toward school and motivation for attending school, learning environment of the math class, use of computer technology, receipt of special services, time spent on homework, importance of grades to the student, school-sponsored activities (sports and extracurricular activities), time spent in reading and outside activities (including television viewing and video games), and use of

the library media center. There are also questions (with parallels on the parent instrument) about close school friends and their parents that are intended, among other uses, to measure aspects of embeddedness in social networks that might be a source of social capital.

The third module of the student questionnaire concerns *plans for the future*. Many elements of the series of life goals questions have been asked since NLS-72. Another question series concerns the persons who have the most influence on the sophomore's plans for the time after high school graduation. Questions are also asked about educational attainment—both the sophomore's perception of parental aspirations for them as well as their personal expectations for highest level of education to be completed. Several items ask about planning for postsecondary education, such as plans for taking the Scholastic Assessment Test (SAT), American College Test (ACT), or other tests, and where students obtain information about various colleges. Other items ask about their desired job after high school (if going directly into the labor force) or job/occupation at age 30, when most cohort members will have completed their postsecondary education and most will have assumed occupational roles.

The section on language use is aimed at students for whom English is not their native language. Items attempt to identify the native language and to address issues of language acquisition, usage, and the extent to which students' limited English skills affect academic achievement, aspirations, and opportunities. These data can be linked to parent questionnaire data on length of residence in the United States and immigration history.

The module on *money and work* provides information to identify the type and amount of work that sophomores are engaged in after school and on weekends. Questions are asked about employment type, hours worked, wages earned, participation in work-based learning programs, how students got their job, and whether the job is related to what they would like to do in the future.

The section on the sophomore's *family* contains questions that will render information about the student's family background and characteristics. Even though redundant with the parent questionnaire, questions are asked about the education and occupation of students' parents. A number of items ask about parental monitoring, as perceived by the student, including checking on homework, limiting of television viewing time, requirements such as chores, limitation of amount of time going out with friends on school nights, and so on. An additional question series gets at the frequency of student-parent discussions on various topics (course selection, grades, college planning, etc.).

The final section of the student questionnaire is a module on beliefs and opinions about self. Included are a number of psychological scales, which have been adapted from PISA:2000. The scales are: (1) instrumental motivation (utility interest); (2) intrinsic interest (specific to mathematics and to English); (3) general control beliefs and expectations concerning the student's capability to perform a task; and (4) self-efficacy (specific to mathematics and to English). A further strand of content concerns peer relations and friends' behaviors, dropout propensities, and values.

2.2.2 Reading and Mathematics Assessments

This section describes the development and format of the tests, the scoring procedures, score descriptions, and summary statistics. It includes a discussion of links (through equating or concordance) with other studies (HS&B, NELS:88, and PISA:2000).

The purpose of the ELS:2002 assessment battery is to provide measures of student achievement in reading and mathematics that can be related to student background variables and educational processes, for individuals and for population subgroups. The reading and mathematics tests must provide accurate measurement of the status of individuals at a given point in time. The mathematics test must provide accurate measurement of their cognitive growth over time. Assessment data in ELS:2002 will be used to study factors that contribute to individual and subgroup differences in achievement.

2.2.2.1 Test Design and Format

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 and were divided into process categories of skill/knowledge, understanding/ comprehension, and problem solving. 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.

The test questions were selected from previous assessments: NELS:88, NAEP, and PISA. Items in both domains were field tested 1 year prior to the 10th-grade survey, and some items were modified based on field test results. Final forms for 10th grade were assembled based on psychometric characteristics and coverage of framework categories. All of the reading questions and about 90 percent of the mathematics questions were presented in multiple-choice format. The 10 percent of math questions that were open ended were scored as right or wrong, with no partial credit awarded.

The 10th-grade tests were administered in two stages. All students received a multiple-choice routing test composed of two separate parts: a 15-question mathematics section, followed by 14 reading questions. The answer sheets were scored by survey administrators, who then assigned each student to a low, middle, or high difficulty second stage form in each subject, depending on the student's number of correct answers in the routing test. The second-stage test forms contained free response as well as multiple-choice items. The two-stage procedure was designed to maximize the accuracy of measurement that could be achieved in a limited amount of testing time, while minimizing floor and ceiling effects (for definitions of floor effects, ceiling effects, and other technical terms, see the glossary in appendix E).

Two of the survey schools were unable to allot enough time for students to participate in the two-stage testing procedure. In these schools, only a single broad-range mathematics form was administered.

Table 1 shows the number of test items in each of the test forms, including Form V, the short mathematics test administered in schools when time was limited. Test scores for a domain were calculated for students who responded to at least 10 items on the routing test and second-stage test, combined.

2.2.2.2 Scoring Procedures

The scores used to describe students' performance on the direct cognitive assessment are broad-based measures that report performance in each domain (mathematics and reading) as a whole. The scores are based on Item Response Theory (IRT), which uses patterns of correct, incorrect, and omitted answers to obtain ability estimates that are comparable across different test forms within a domain.[11] In estimating a student's ability, IRT also accounts for each test question's difficulty, discriminating ability, and a guessing factor.

Table 1. Number of items in each ELS:2002 test form for assessing achievement in mathematics and reading: 2002

Form	Mathematics	Reading
Routing test	15	14
Second stage tests		
Form X (low difficulty)	25	16
Form Y (middle difficulty)	27	17
Form Z (high difficulty)	27	15
Form V (single stage: limited time, mathematics only)	23	†

†Not applicable.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

IRT has several advantages over raw number-right scoring. By using the overall pattern of right and wrong responses to estimate ability, IRT can compensate for the possibility of a low-ability student guessing several difficult items correctly. If answers on several easy items are wrong, a correct difficult item is assumed, in effect, to have been guessed. Omitted items are also less likely to cause distortion of scores, as long as enough items have been answered right and wrong to establish a consistent pattern. Unlike raw number-right scoring, which necessarily treats omitted items as if they had been answered incorrectly, IRT procedures use the pattern of responses to estimate the probability of correct responses for all test questions. Finally, IRT scoring makes it possible to compare scores obtained from test forms of different difficulty. The common items present in the routing test and in overlapping second-stage forms allow test scores to be placed on the same scale. Looking ahead to plans for the ELS:2002 first follow-up survey, IRT procedures will be used to estimate longitudinal gains in achievement over time by using common items present in both the 10th- and 12th-grade forms.

2.2.2.3 Score Descriptions and Summary Statistics

Several different types of scores that can be used to describe students' performance on the cognitive assessment are described in detail below. IRT-estimated number right scores measure students' performance on the whole item pool for each domain. NELS:88-equated number right scores estimate how a student would have performed on the 1992 reading and

[11] For an account of Item Response Theory, see Hambleton (1989) or Hambleton (1991).

mathematics scales of NELS:88. Standardized scores (T-scores) report students' performance relative to their peers. Quartile scores divide the estimated population distributions for convenience in analyzing relationships of cognitive skills with other variables. NELS:88-equated proficiency probabilities estimate the probability that a given student would have demonstrated proficiency for each of the three reading and five mathematics levels defined for the NELS:88 survey in 1992.[12]

The database also reports scores for ELS:2002 participants on the mathematics score scale used for NELS:88 10th graders in 199013 and on the PISA:2000 reading scale.

IRT-estimated number right. Scores for mathematics and reading are estimates of the number of items students would have answered correctly if they had responded to all of the 73 questions in the math item pool (i.e., all items that appeared on any of the first- and second-stage mathematics forms) and all 51 questions in the reading item pool. The ability estimates and item parameters derived from the IRT calibration can be used to calculate each student's probability of a correct answer for each of the items in the pool. These probabilities are summed to produce the IRT-estimated number right scores. These scores are not integers because they are sums of probabilities, not counts of right and wrong answers. (Note that scores for different subject areas are not comparable to each other because they are based on different numbers of test questions and on content that is not necessarily equivalent in difficulty. Thus, it would not be correct to assume that a student is doing better in reading than in mathematics because his or her IRT-estimated number right score in reading is higher.)

Table 2 shows variable names, descriptions, and summary statistics for the IRT-estimated number right score. The reliabilities shown in the table are a function of the variance of repeated estimates of the IRT ability parameter and apply to all scores derived from the IRT estimation, including the standardized and quartile scores.

Table 2. Item Response Theory (IRT)-estimated number right scores from ELS:2002 mathematics and reading assessments: 2002

Variable name	Description	Range	Weighted mean	Weighted standard deviation	Reliability
BYTXMIRR	Mathematics IRT-estimated number right	0–73	37.4	12.3	0.92
BYTXRIRR	Reading IRT-estimated number right	0–51	29.4	9.9	0.86

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

[12] For further information on the NELS:88 proficiency levels, see Rock and Pollack (1995), *Psychometric Report for the NELS:88 Base Year Through Second Follow-Up* (NCES 95–382). For examples of the use of the NELS:88-equated probability proficiency scores in the context of cross-sectional estimation of status in ELS:2002, see chapter 5 of Ingels, Burns, Chen, Cataldi, and Charleston (2004), *A Profile of the American High School Sophomore in 2002* (NCES 2004–396). For examples of longitudinal use of the probability of proficiency scores (in NELS:88), see chapter 4 of Scott, Rock, Pollack, and Ingels (1995), *Two Years Later: Cognitive Gains and School Transitions of NELS:88 Eighth Graders* (NCES 95–436).

[13] The 1990 NELS:88 mathematics scale (58 items) is documented in chapter VI of Ingels, Scott, Rock, Pollack, and Rasinski (1994), *NELS:88 First Follow-up Final Technical Report* (NCES 94–632). The 1992 scales (81 items in mathematics, 54 in reading) are documented in Rock and Pollack (1995), *Psychometric Report for the NELS:88 Base Year Through Second Follow-Up* (NCES 95–382).

Standardized scores (T-scores). T-scores 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 other groups. It does not represent mastery of a particular set of skills, only that the subgroup's mastery level is greater than a comparison group. In other words, T-scores provide information on status compared to students' peers, while the IRT-estimated number-right scores represent status with respect to achievement on a particular criterion set of test items. The T-scores can only provide an indicator of the extent to which an individual or a subgroup ranks higher or lower than the national average.

The standardized scores reported in the database are transformations of the IRT theta (ability) estimates, rescaled to a mean of 50 and standard deviation of 10. See table 3 for variable names, descriptions, and ranges for the standardized (T) scores. Weighted means and standard deviations are not included in this table because, by definition, the scores are computed such that the weighted mean (population estimate) is 50.0 and standard deviation 10.0 for each score. The composite score is the average of the mathematics and reading standardized scores, re-standardized to a national mean of 50.0 and standard deviation of 10.0. A few students had scores for only the mathematics test or reading test, but not both. For these students, the composite is based on the single score that was available.

Table 3. Standardized scores (T-scores) from ELS:2002 mathematics and reading assessments: 2002

Variable name	Description	Range
BYTXMSTD	Mathematics standardized score (T-score)	10–90
BYTXRSTD	Reading standardized score (T-score)	10–90
BYTXCSTD	Composite mathematics + reading standardized score (T-score)	10–90

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Quartile scores divide the weighted (population estimate) achievement distributions into four equal groups, based on mathematics, reading, and mathematics + reading composite scores. Quartile 1 corresponds to the lowest achieving quarter of the population, quartile 4 to the highest. Table 4 contains variable names, descriptions, and ranges for the quartile scores.

Table 4. Quartile scores from ELS:2002 mathematics and reading assessments: 2002

Variable name	Description	Range
BYTXMQU	Mathematics quartile	1–4
BYTXRQU	Reading quartile	1–4
BYTXCQU	Composite mathematics + reading quartile	1–4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

2.2.2.4 Links to Other Surveys

Scores for ELS:2002 are reported on scales that permit comparisons with reading and mathematics data for NELS:88 10th graders, and with PISA:2000 reading results for 15-year-olds. The link to the NELS:88 scales represents a "true" equating. This means that the tests may be considered interchangeable or, in other words, a score on one exam should be equivalent to a score on the other exam. Several conditions must be met for equating two tests. Most

importantly, the tests must measure the same content. Similarity of format, length, reliability, and subgroup performance also support the interpretation of interchangeable scores. The differences between ELS:2002 and PISA, described below, were sufficiently important that the PISA scale scores represent a concordance, or a link based on population distributions, rather than equivalent or interchangeable scores.

NELS:88-equated Scores. 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. (The ELS:2002 data for seven of the reading items and six of the math items did not fit the NELS:88 IRT parameters, so these items were not treated as common items for the purpose of equating.) 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.

The NELS:88-equated IRT-estimated number right scores for reading and mathematics are estimates of the number of items students would have answered correctly had they taken the NELS:88 exam and responded to all items in the mathematics items pool or to all items in the reading item pool, respectively. The NELS:88 item pool contained 81 mathematics items and 54 reading items in all test forms administered in grades 8, 10, and 12. An additional mathematics score, based on the 58 NELS:88 items that appeared in the grades 8 and 10 mathematics forms, is also provided. The 1990-equated mathematics score (BYNELS0M) was generated for the specific purpose of supporting comparisons with HS&B in 1980 and NELS:88 in 1990 (HS&B results were placed on the NELS:88 58-item mathematics scale). Table 5 shows reading and mathematics scores for ELS students, reported on the various NELS:88 score scales.

Table 5. ELS:2002 Item Response Theory (IRT) NELS-equated estimated number right scores: 2002

Variable name	Description	Range	Weighted mean	Weighted standard deviation	Reliability
BYNELS2R	Reading—NELS-equated estimated number right (1992 scale)	0–54	29.2	9.5	0.87
BYNELS2M	Mathematics—NELS-equated estimated number right (1992 scale)	0–81	44.4	13.7	0.92
BYNELS0M	Mathematics—NELS-equated estimated number right based on 58 items (1990 scale)	0–58	37.6	11.4	0.92

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

The criterion-referenced NELS-equated proficiency probability scores are based on clusters of items that mark different levels on the reading and mathematics scales. Clusters of four items each were identified in the NELS:88 tests that marked three hierarchical levels in reading and five in mathematics.

- *Reading levels:*

    1. Simple reading comprehension, including reproduction of detail and/or the author's main thought.

2. Simple inferences beyond the author's main thought, and/or understanding and evaluating abstract concepts.
3. Complex inferences or evaluative judgments requiring multiple sources of information.

- *Mathematics levels:*

1. Simple arithmetical operations on whole numbers.
2. Simple operations with decimals, fractions, powers, and roots.
3. Simple problem solving requiring the understanding of low-level mathematical concepts.
4. Understanding of intermediate-level mathematical concepts and/or multi-step solutions to word problems.
5. Complex multi-step word problems and/or advanced mathematical material.

The proficiency levels are hierarchical in the sense that mastery of a higher level typically implies proficiency at lower levels. In NELS:88, students were judged to be proficient if three of the four items in a cluster were answered correctly. The NELS:88-equated proficiency probabilities were computed using IRT-estimated item parameters calibrated in NELS:88. Each proficiency probability represents the probability that a student would pass a given proficiency level defined as above in the NELS:88 sample.

Table 6 shows variable names, descriptions, and summary statistics for the NELS-equated proficiency probability scores.

Table 6. Reading and mathematics probability of NELS-equated proficiency scores: 2002

Variable name	Description	Range	Weighted mean	Weighted standard deviation
BYTX1RPP	Reading—level 1	0–1	0.89	0.26
BYTX2RPP	Reading—level 2	0–1	0.46	0.40
BYTX3RPP	Reading—level 3	0–1	0.08	0.21
BYTX1MPP	Mathematics—level 1	0–1	0.92	0.20
BYTX2MPP	Mathematics—level 2	0–1	0.67	0.42
BYTX3MPP	Mathematics—level 3	0–1	0.46	0.46
BYTX4MPP	Mathematics—level 4	0–1	0.21	0.33
BYTX5MPP	Mathematics—level 5	0–1	0.01	0.07

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

PISA Concordance. The ELS:2002 and PISA reading tests did not share enough items to permit common-item equating, so score scales were linked by means of *equipercentile equating*. If two exams measuring the same construct are given to two samples from the same population, the score corresponding to a certain percentile on one exam may be considered to be equivalent to the score on the other exam that represents the same percentile of the population. ELS:2002 and PISA test instruments, scoring methods, and populations differed in several respects that impact the equating procedures and interpretation of linked scores.

PISA reading items tended to focus on reading applications, including diagrams and graphs. While some PISA reading materials were incorporated in the ELS:2002 tests, other passages were taken from NELS:88, which consisted entirely of text. All ELS:2002 students received approximately 30 reading items, while PISA takers might have had a wide range of numbers of items. Some PISA test booklets had very few reading items, others many more. Scores based on very few items would be expected to have relatively low reliability. The scoring methods employed also differed: ELS:2002 based scores on a three-parameter IRT model, while PISA used one-parameter IRT. PISA scoring treated omitted items as wrong for some purposes; ELS:2002 scoring treated them as unanswered or not administered.

The most important difference between PISA and ELS:2002 is the definition of the population sampled in each case. Equipercentile equating assumes that the two samples being equated come from the same population. However, important differences exist between PISA and ELS:2002 (see table 7 below). The PISA population was based on age (students born in 1984), while ELS:2002's population was based on grade (high school sophomores). While the spring term administration dates for PISA and ELS:2002 overlapped, the range of PISA dates was later in the school year, suggesting the possibility of higher scores due to additional weeks or months of schooling.

Table 7. ELS:2002 and Program for International Student Assessment:2000 (PISA:2000) samples: 2002

ELS:2002 sample	PISA:2000 sample
10th graders only	Different grades
Different ages; modal age=15	Ages 15.25–16.25 years
Testing began in January 2002	Testing began in April 2000
14,543 tested	3,700 tested

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Because of these differences, subsamples of each of the groups were used to compute equivalent percentiles (see table 8). Transformations were computed based on the 10th graders from each survey who were within a specified range of ages and testing dates. The resulting transformation was then applied to all ELS:2002 students. To make the PISA sample more nearly equivalent to the ELS:2002 sample, only PISA 10th graders were used in the equating subsample. To make the ELS:2002 sample more nearly equivalent to the PISA sample, only ELS:2002 students between the ages of 15.25 years and 16.25 years (the approximate age range for PISA examinees) were used. ELS:2002 students who were tested before March 15 or after May 31 were deleted from the equating sample. The restricted samples were intended to be representative of 10th graders between the ages of 15.25 and 16.25 years.

Table 8. ELS:2002 and Program for International Student Assessment:2000 (PISA:2000) equating sample: 2002

ELS:2002 equating sample	PISA:2000 equating sample
10th graders only	10th graders only
15.25- to 16.25-year-olds	15.25- to 16.25-year-olds
Exams offered from March 15 to May 31	Exams offered from April 1 to May 31
Equating sample N=2,694	Equating sample N=2,207

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

After these factors had been controlled, the ELS:2002 and PISA subsamples still did not represent exactly the same population. The PISA survey was carried out 2 years earlier than ELS:2002. Demographic trends or cohort differences could conceivably affect distributions of achievement even in so short a time. Differences in weighted population estimates were reviewed for the relevantly overlapping samples for each survey. The ELS:2002 population estimate was 15 percent greater than the PISA estimate (2,016,756 to 1,760,892). Percentages of racial/ethnic groups were quite similar for several but not all groups, with variations from 0.1 percent to 5.1 percent. It is impossible to tell whether overall and racial/ethnic differences are due to sampling variability, differences in racial/ethnic identification methods, or to other differences, such as the higher rate of missing race/ethnicity data in PISA (in PISA, race/ethnicity identification is available for 93.4 percent of the overall sample; in ELS:2002, race/ethnicity identification was gathered for 99.98 percent of the sample, then imputed for the missing 0.02 percent) (table 9).

Table 9. Comparative statistics for full-sample Program for International Student Assessment:2000 (PISA:2000) and ELS:2002 base year: 2002

Race	Weighted frequency distribution	Weighted percent
PISA:2000		
White	1,725,766	55.3
American Indian	93,471	3.0
Black	407,593	13.1
Multiracial	48,088	1.5
Hawaiian	10,847	0.3
Asian	105,183	3.4
Hispanic	523,996	16.8
Missing	207,040	6.6
Total	3,121,874	100.0
ELS:2002		
White, non Hispanic	2,077,826	60.4
American Indian or Alaska Native	32,284	0.9
Black or African American, non-Hispanic	495,642	14.4
Multiracial, non-Hispanic	148,232	4.3
Native Hawaiian or Other Pacific Islander	8,244	0.2
Asian, non-Hispanic	130,050	3.8
Hispanic or Latino	547,211	16.0
Missing	0	0.0
Total	3,439,490	100.0

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

The equipercentile equating was carried out using three-moment smoothing of the weighted frequency distributions. Plots of the equipercentile-equated scores showed extreme deviations in the tails of the distribution from a trend line based on linear approximation. These deviations are probably due to the methodology employed in PISA scoring: the PISA scores are transformations of normally distributed IRT ability estimates, which, if no shrinkage is imposed, tend to have long tails. The ELS:2002 scores, which are sums of probabilities, do not. As a result, the equipercentile conversion becomes distorted in the tails of the distributions. Throughout most of the score range, a quarter point difference in ELS:2002 reading scale corresponds to a difference of 2 to 3 points in the PISA metric. But, in the extreme tails of the distribution, a quarter point difference in ELS:2002 reading score corresponds to a difference of 5 to 10 points or more in the PISA metric. For this reason, the equipercentile equating was carried out without the data in the top and bottom tails of each distribution. Then the equipercentile transformation was used to link the scores for the middle 90 percent of the students, and the remaining scores were linked based on the linear approximation of the equating transformation. The cut-off points for using equipercentile versus linear transformation were selected such that the ELS:2002 to PISA link would be monotonic. Table 10 shows the linking methods for implementing PISA:2000 reading scales in ELS:2002.

Table 10. Linking methods for implementing Program for International Student Assessment:2000 (PISA:2000) reading scales in ELS:2002: 2002

ELS:2002 scale score range	Equating method	Weighted percent of data
10.00–13.50	Linear approximation	5.3
13.50–45.00	Equipercentile transformation	90.4
45.00–49.25	Linear approximation	4.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Data users should keep in mind that the differences between the ELS:2002 and PISA:2000 tests, scoring methods, and populations mean that the link reported here cannot be considered to be a true equating. Although procedures were carried out to compensate for population differences and scoring methods, no claim is made that the scores may be treated as equivalent. It is more appropriate to refer to this link as a concordance: the PISA-scale score represents the score level achieved by students of the same percentile rank in two populations that were matched as closely as was possible given the differences described above.

Choosing the Appropriate Score for Analysis. The IRT-estimated number right, standardized scores (T-scores), proficiency, and quartile scores are all derived from the IRT model, and are based on all of the student's responses to a subject area assessment. That is, the pattern of right and wrong answers, as well as the characteristics of the assessment items themselves, are used to estimate a point on an ability continuum, and this ability estimate, theta, then provides the basis for criterion-referenced and norm-referenced scores. The choice of the most appropriate score for analysis purposes should be driven by the context in which it is to be used.

The *IRT-estimated number right scores* are overall, criterion-referenced measures of status at a point in time. The criterion is the set of skills defined by the framework and represented by the assessment item pool. These scores are useful in identifying cross-sectional differences among subgroups in overall achievement level. They provide a summary measure of

achievement useful for correlational analysis with status variables, such as demographics, school type, or behavioral measures, and may be used in multivariate models as well.

The standardized scores (T-scores) are also overall measures of status at a point in time, but they are norm-referenced rather than criterion-referenced. They do not answer the question, "What skills do students have?" but rather, "How do they compare with their peers?" The transformation to a familiar metric with a mean of 50 and standard deviation of 10 facilitates comparisons in standard deviation units. For example, an individual with a T-score of 65 (or a subgroup with a mean of 65) has demonstrated achievement one-and-one-half standard deviations above the national average for 10th graders, while a score of 45 would correspond to half a standard deviation below the norm. These numbers do not indicate whether students have mastered a particular body of material, but rather what their standing is relative to others. Unlike the IRT-estimated number right scores, the standardized scores may be used to compare reading and mathematics achievement. For example, one might find that an individual or subgroup excels in math achievement relative to others, but lags behind in reading.

Quartile scores are convenient normative scores for the user who wishes to focus on analysis of background or process variables separately for students at different achievement levels. For example, one might want to compare the school experiences or educational aspirations of students in the lowest reading quartile with those of students in the highest quartile group.

Probability of proficiency scores are criterion-referenced scores that may be used in a number of ways. They are continuous scores that, because they are expressed on the NELS:88 scale, may be used for cross-cohort measurement (i.e., comparing the achievement of NELS:88 and ELS:2002 sophomores in reading and mathematics). They may also be used cross-sectionally to locate the achievement of ELS:2002 sample members and subgroups at various behaviorally defined skill levels. Because their range is 0 to 1, their means can also be expressed in percentage form (e.g., one could say that 20 percent of some given group is proficient in mathematics at level 3, simple problem solving). When mathematics data from the first follow-up (2004) become available, the proficiency scores can be used to measure gain. The proficiency probabilities are particularly appropriate for relating specific processes to changes that occur at different points along the score scale. For example, two groups may have similar gains, but for one group, gain may take place at an upper skill level, and for another, at a lower skill level. For those who gain at the higher skill level, there may be an association between their gains and curriculum exposure, such as taking advanced mathematics classes.

2.3 Parent Questionnaire

The parent questionnaire was to be completed by the parent or guardian most familiar with the sophomore's school situation and experience. Guided by this definition of the preferred respondent, the parent survey respondent was self-selected. Only one parent survey is planned for ELS:2002, the survey that was conducted in the base year.

The parent questionnaire was available in both English and Spanish. Both a hardcopy version and an electronic version for computer-assisted telephone interviews (CATI)[14] were produced. The parent questionnaire addressed the following five topic areas: (1) family background; (2) their child's school life; (3) their child's family life; (4) their opinions about their child's school; and (5) their aspirations and plans for their child's future.

The *family background* module of the questionnaire elicits information about family composition and structure, parent age, education, occupation, religious, and racial/ethnic background. Other questions provide information about immigration patterns and languages spoken in the home.

The section on the sophomore's school life elicits information on the child's educational history; for example, whether the child has been served by Head Start, attended kindergarten, or was held back one or more grades or changed schools. The school life module also asks about the parent's perception of the child's disability status, health, and behavioral problems. Questions are asked as well about reasons for contacts between the parents and the school (both contacts initiated by the parent, and contacts initiated by the school), and about the parents' involvement with the school. Parental monitoring is another topic covered in this module, with questions about checking homework, curfews, and discussions of report cards. Questions are asked as well about the frequency of different kinds of discussions with the sophomore, including planning for higher education and the job market.

The focus of another parent questionnaire module is the sophomore's family life. Here further parental monitoring questions include items on enforced norms about household chores and watching television and the frequency with which parents and the student share meals and participate in various activities together. Questions cover the sophomore's computer access and use, but also computer use by the parent to communicate with or obtain information from the school. "Social capital," with its notion that individuals can access such social resources as networks of information, help, and social support, is measured by two questions in the parent questionnaire and by parallel questions in the student questionnaire. These questions try to learn more about whether there is a functional community linking families to a school. One of the social capital questions asks for information about three of the sophomore's close friends, and the other asks about the parents' interactions with parents of the sophomore's friends. Other questions ask about the community in which the family lives.

Another module elicits the parent's opinions about the sophomore's school, including whether the schoolwork is intellectually challenging; whether the school setting is physically safe, free of drugs, and so on; and about the parent's level of satisfaction with the education that the student has received so far. Questions about the sophomore's future include items about the parents' aspired-for level of educational attainment for their child, their expectations for educational attainment, their financial planning for college, and their income.

[14] The approach to parent telephone interviews in ELS:2002 differed from that followed in NELS:88. In NELS:88, to minimize the possibility of mode of administration effects, the parent was asked to read along in the hardcopy questionnaire as the questions were read over the telephone. The interview was not computer assisted. In ELS:2002, the decision was made to take advantage of the logical consistency editing and other features of CATI, and considerable effort was made to constrain the hardcopy questionnaire to items and formats compatible with a CATI administration. ELS:2002 parents were not interviewed over the telephone with the hardcopy questionnaire in hand. This difference accounts for some differences between the NELS:88 and ELS:2002 parent survey instruments.

2.4 Teacher Questionnaire

The teacher questionnaire was to be completed by the English teacher and the mathematics teacher of each ELS:2002 sophomore. The teacher questionnaire was designed to illuminate questions of the quality, equality, and diversity of educational opportunity by obtaining information in two content areas:

- *Teacher Evaluations of Students*. The teacher's assessment of the student's school-related behavior and academic performance and educational and career plans and goals. Respondents complete this section with respect to the sample members they instructed in a particular subject.
- *Teacher Background*. Information about the teacher's background and activities (e.g., academic training, subject areas of instruction, years of teaching experience, and participation in professional growth activities).

2.4.1 Teacher Evaluations of ELS:2002 Students

Teacher evaluations are elicited along a number of dimensions of student motivation and performance. Teachers are asked to rate how hard the student works for good grades in the class; whether homework assignments are typically completed; and how often the student is absent, tardy, attentive, or disruptive. Other questions inquire into communications with the student's parents and degree of parental involvement. Teachers are asked how far in school they expect the student to get. English teachers are asked to rate the student's compositional skills.

2.4.2 Teacher Background

The *teacher background* section inquires into the teacher's social and educational background, professional experience, on-the-job training, and social networks. Items collected include basic teacher demographics (sex, race, date of birth), years in teaching and at the school, full-time versus part-time and certification status, academic degrees, field of study, job satisfaction, and attributions of student success. New items have been added about the teacher's experience with computers and other aspects of technology.

The teacher questionnaire was designed to provide data that can be used in analyzing influences on student sample members. The design of the component does not provide a stand-alone analysis sample of teachers—either of teachers in the nation, or of teachers in the school.

2.5 School Administrator Questionnaire

The school administrator questionnaire collects information on the school in six areas: school characteristics, student characteristics, teaching staff characteristics, school policies and programs, technology, and school governance and climate. Data gathered in the school administrator questionnaire can be merged with data from the student and teacher questionnaires and the student cognitive test battery. This linkage of the data will allow researchers to examine to what degree disparities in educational aspirations, expectations, and outcomes of various student populations are accounted for by differences in the schools that these students attend.

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 ELS:2002 school sample is nationally representative and can stand alone as a basis for generalizing to the nation's regular high schools with sophomores in the 2001–02 school year. (While, owing to the births and deaths of schools over a 2-year period, the school sample in 2004 will no longer be strictly nationally representative, ELS:2002 will return to the same schools 2 years hence. It is therefore possible to look at the base year school sample longitudinally—the nation's 2002 high schools, 2 years later.) Indeed, since ELS:2002 comprises a panel of schools as well as students, school-level analyses can be conducted in which the student data are aggregated to represent the sophomore class of the school, and outcomes measured 2 years later for those students who remain in the school.

It should be noted that, in many cases, the facsimiles of the school administrator questionnaire show that items were asked in continuous form, but in the public-use codebook, these items appeared only in categorical form. Considerable recoding (including top and bottom coding) has been implemented in the school-level data as a protection against deductive disclosure of school identities. Researchers who require the more fine-grained versions of these variables should apply to NCES for a special license to use restricted data.

2.5.1 School Characteristics

The *school characteristics* section of the school administrator questionnaire collects contextual information such as school type; grade levels in the school; the school calendar; length of school year, day, and class periods; and availability of and student participation in courses, programs, and services.

2.5.2 Student Characteristics

The *student characteristics* module elicits information on characteristics of the school's sophomore class, for example, the percentage of sophomores who are limited English proficient (LEP) and the percentage who receive free or reduced-price lunch. It should be noted that additional characteristics of sophomores in the school are made available on the data file from nonquestionnaire sources such as the sampling lists (sophomore enrollment), or from the NCES Common Core of Data (CCD) and the Private School Survey (PSS) (e.g., the racial/ethnic percentages for students at the school, though this specific datum is available only on the restricted-use file).

2.5.3 Teaching Staff Characteristics

The *teaching staff characteristics* section of the school administrator questionnaire collects the number of full-time and part-time teachers overall and by subject area, teacher compensation, teacher evaluation standards and methods, and teacher rewards.

2.5.4 School Policies and Programs

The module on *school policies and programs* addresses two areas of school policy: competency tests for high school graduation and school safety and crime. Schools that administer competency tests are asked to indicate the grades in which the tests are given; whether the exams are a requirement of the state, district, or the school; the subject areas covered

on the tests; and the percentage who fail the exams and the recourse for these students. School administrators are also questioned about security measures taken by the school and efforts to involve parents in school discipline and safety.

2.5.5 Technology

A number of questions are asked about school technology. These questions elicit information about teacher access to various technologies for instructional use; teacher use of computers for various purposes (professional development, communication with parents, instruction); administrators' computer use; and technology training for teachers.

2.5.6 School Governance and Climate

The final module of the school administrator questionnaire concerns *school governance and climate*. This section is to be completed by the school's chief administrator and cannot be delegated to subordinate staff. The module addresses the decision-making authority of the school principal and the principal's accountability, the incidence of problem behaviors of students, the quality of relations with individuals and organizations in the community, the condition of facilities, and the school ethos or climate.

It should be noted that a subset of the school administrator questionnaire items were also asked of nonresponding schools, either directly or through their districts. Data from this abbreviated version of the questionnaire, the school characteristics questionnaire, are not available on the ELS:2002 data files. However, the information was used to conduct a bias analysis in which characteristics of responding and nonresponding schools were compared.

2.6 Library Media Center Questionnaire

A library media center questionnaire was not a feature of ELS:2002's predecessor studies. The addition of this component will permit investigation of the role of integration between library media resources and curriculum and classroom instruction in promoting effective learning. 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; 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 also supplies 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, data from the questionnaire can be used at the school level to generalize to libraries within all regular high schools with 10th grades in the United States in the 2001–02 school year. The ELS:2002 library media center questionnaire is largely an abridgment of the school library media center questionnaire used in SASS:2000.

2.7 School Facilities Checklist

Instrumentation for 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 from the checklist 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.

Chapter 3
Sample Design, Weighting, Design Effects, and Data Quality

3.1 Introduction

Chapter 3 describes the base year sample design, unit and item nonresponse bias analysis, imputation, weighting, standard errors and design effects, and disclosure analysis. This section provides an overview of each of these subjects, and the details are provided in later sections of chapter 3.

3.1.1 Sample Design

The sample design for the Education Longitudinal Study of 2002 (ELS:2002) is similar in many respects to the designs used in the three prior studies of the National Education Longitudinal Studies Program, the National Longitudinal Study of the High School Class of 1972 (NLS-72), the High School and Beyond (HS&B) longitudinal study, and the National Education Longitudinal Study of 1988 (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 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 (PPS), and school contacting resulted in 1,221 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. A full discussion of the sample design and response rates is presented in chapters 3 and 4. 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.

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. Schools and students are intended as the study's basic units of analysis. 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. Student-level data consist of student questionnaire and assessment data, and reports from students' teachers and parents. (School-level data, however, can also be reported at the student level and serve as contextual data for students.)

3.1.2 Schools

The target population of schools for the full-scale ELS:2002 study consisted of regular public schools, including State Education Agency schools and charter schools, and Catholic and other private schools that contain 10th grades and are 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., data collection occurred for students in the school).[15] Of the 1,268 sampled schools, 1,221 were eligible. Some 752 of the 1,221 eligible schools responded for a 67.8 percent (weighted[16]) response rate.

A subset of most but not all responding schools also completed a school administrator questionnaire and a library or media center questionnaire (98.5 percent and 95.9 percent weighted response rates, respectively; see section 4.9). Most nonresponding schools or their districts provided some basic information about school characteristics, so that the differences between responding and nonresponding schools could be better understood, analyzed, and adjusted (see section 3.2.6). Additionally, the RTI field staff completed a facilities checklist for each responding school (100 percent response rate; see section 4.9).

3.1.3 Students

The target population of students for the full-scale ELS:2002 consisted of spring-term 10th graders 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.

The ELS:2002 survey instruments comprised two assessments (reading and mathematics) and a student questionnaire. Participation in ELS:2002 was defined by questionnaire completion. Some 87.3 percent (weighted[17]) of eligible selected students participated by completing the student questionnaire. While in general students were asked to complete the assessment battery in addition to the questionnaire, there were some cases in which a student completed the questionnaire but did not complete the assessments.

The following guidelines were provided to schools to assist them in determining whether students would be able to complete the ELS:2002 survey instruments. First, for students whose native language was not English and whose English language proficiency was limited, such students were deemed to be able to participate if either (a) the student had received academic instruction primarily in English for at least 3 years or (b) in the school's judgment, it was felt that the student could meaningfully respond to the questionnaire or validly be assessed. Second, for students whose mental or physical disabilities constituted a potential barrier to participation, the following guidelines were offered. If a student's individualized education program (IEP) indicated that the student should not be tested through standardized pencil-and-paper assessments, the student was not asked to complete the ELS:2002 assessment battery. (The

[15] One eligible school had no eligible students selected in the sample. This school was considered a responding school.
[16] The weight for school response rate computation was the school weight prior to nonresponse adjustment, i.e., WT1*WT2*WT3, as described in section 3.4.1.
[17] The weight for student response rate computation was the student weight prior to nonresponse adjustment, i.e., final weight*WT6, as described in section 3.4.2.

school had to make a further determination as to whether such a student could complete the questionnaire.) If the student's IEP indicated that the student could be tested with accommodations, the following accommodations were acceptable, if it was possible (in practical terms) to implement them:

- extra time;
- split session;
- instructions in sign language for hearing-impaired students;
- enlarged questionnaire or assessment booklet for the visually impaired;
- one-on-one session (if the student could not participate in group settings); and
- completion of the student questionnaire through personal interview. This option was available to students unable to read visually based text (e.g., blind students), even though they could not complete the assessments. Likewise, students unable to effectively read and respond to the questionnaire because of a learning disability could be administered the questionnaire in a personal interview.

Students who could not (by virtue of limited English proficiency or physical or mental disability) complete the ELS:2002 survey instruments (including the questionnaire as well as the tests) were part of the expanded sample of 2002 sophomores who will be followed in the study and whose eligibility status will be reassessed 2 years hence. There were 163 such students. These students appear only on the base year (and will appear in subsequent) restricted-use files and provide usable information to the extent that their school enrollment status is ascertained in the base year (and then again in the first follow-up). To obtain additional information about their home background and school experiences, contextual data (base year parent survey, base year teacher data) were used to classify students in computer-assisted telephone interviewing (CATI).

3.1.4 Parents

Parent data have been collected to support analyses at the student level. Conceptually, the universe of parents of 10th-grade students comprised all parents or legal guardians of 10th-grade students in spring 2002. Once the full sample of 10th graders was selected, the parent or guardian who was best informed about the child's educational activities was asked to complete an ELS:2002 parent questionnaire. The selection of parents thus did not require the construction of a formal universe list or frame. It is important to remember that the student is the central unit of analysis and that parent reports were collected to provide contextual data for students.

Parents were eligible if they fit the above definition and were respondents if they completed the parent questionnaire. The overall weighted parent coverage rate was 87.4 percent, conditional on student response (see chapter 4 for more detailed response and coverage rates).

3.1.5 Teachers

Teacher data also have been collected to support analyses at the student level. All full- and part-time teachers who were teaching 10th graders in mathematics or English/language arts in spring 2002 were included in the ELS:2002 universe of 10th-grade teachers. The actual sample was restricted to teachers who were providing instruction in one of the two subjects to the

full (expanded) sample of 10th-grade students within the sampled schools. Thus, as for parents, there was no need to construct a formal universe list of 10th-grade mathematics and English teachers prior to their selection. Once students were selected within schools, their teachers of the assigned subject pairs were identified and asked to participate in the study. It is important to remember that the student is the central unit of analysis and that teacher reports were collected to provide contextual data for students.

Teachers were eligible if they fit the above definition and taught the student at the sample school and were respondents if they completed the teacher questionnaire for at least one student. The responding teachers covered 91.6 percent of the responding students, that is, the weighted coverage rate was 91.6 percent (see chapter 4 for more detailed response and coverage rates).

3.1.6 Nonresponse Bias Analysis

The overall weighted school participation rate was 67.8 percent. The overall weighted student response rate was 87.3 percent, although the response rate for certain domains was below 85 percent. School and student unit nonresponse bias analyses were performed. The bias due to nonresponse prior to computing weights and after computing weights was estimated based on the data collected from both respondents and nonrespondents, as well as frame data. An item nonresponse bias analysis was also performed for all questionnaire variables in which response fell below 85 percent. Details of the bias analyses are given in 3.2.6.

3.1.7 Imputation

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. These variables were chosen because they are key variables used in standard reporting and cross-sectional estimation. Imputation of missing values for the ability estimates provided complete information for the various test scores derived from those estimates, including both normative and criterion-referenced scores. The imputations were performed primarily to reduce the bias of survey estimates caused by missing data. The imputed data also made the data more complete and easier to analyze. Most of the variables were imputed using a weighted hot-deck procedure.[18] The order of imputation addressed problems of multivariate association by using a series of univariate models fitted sequentially such that variables modeled earlier in the hierarchy had a chance to be included in the covariate set for subsequent models. Additionally, multiple imputations were used for a few variables, including test scores. Imputation is discussed in detail in section 3.3.

3.1.8 Weighting

The general purpose of the weighting scheme was to compensate for unequal probabilities of selection of schools and students into the base year sample and to adjust for the fact that not all schools and students selected into the sample actually participated. Three sets of weights were computed: a school weight, a weight for student questionnaire completion, and a contextual data weight for the "expanded" sample of questionnaire-eligible and questionnaire-ineligible students. Schools and students were adjusted for nonresponse, and these adjustments

[18] See Cox (1980).

were designed to significantly reduce or eliminate nonresponse bias for data elements known for most respondents and nonrespondents. In addition, school weights were poststratified to known population totals. Weighting is discussed in detail in section 3.4.

3.1.9 Standard Errors and Design Effects

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 departures 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. Appendix K presents standard errors and design effects for 30 means and proportions based on the ELS:2002 student, parent, and school data for the sample (as a whole and for selected subgroups).

3.1.10 Disclosure Analysis

Because of the paramount importance of protecting the confidentiality of NCES data containing information about specific individuals, ELS:2002 data were subject to various procedures to minimize disclosure. As a first step, all ELS:2002 data files (school, student, teacher, and parent) were reviewed to identify high-risk variables. As a second step, a technique called "data swapping" was carried out, both for school-level data and for student-level data (student, parent, and teacher). As a final step, the ELS:2002 data underwent a disclosure risk analysis. In this analysis, school characteristics information available on the data files was compared to information on publicly available universe files of schools.

3.2 Base Year Sample Design

3.2.1 Sampling Frame

The preliminary 1999–2000 CCD and the provisional 1999–2000 PSS data files of public and private schools, respectively, were used as the sampling frames.

The survey population for the full-scale ELS:2002 consisted of spring-term 10th graders in 2002 who were enrolled in school in the United States (50 states and the District of Columbia) in regular public schools, including State Department of Education schools and charter schools, and in Catholic and other private schools.

These types of schools were excluded from the school sampling frame:

- *Schools With No 10th Grade.* The low grade and high grade indicators were used to identify such schools. However, several hundred schools had grade levels that did not include a 10th grade but did have a positive 10th-grade enrollment. Some schools had a grade span that included 10th grade but that had zero 10th-grade enrollments.

These schools were included as long as the school name did not indicate that the school was an elementary, middle, or junior high school.

- *Schools With No Enrollment.* A school with no enrollment indicated that the school did not directly enroll students. Students at such schools were counted with their "home" school, and they are included in the student population.
- *Ungraded Schools.* If the low grade and high grade indicators were both 'UG' or '00,' the school was classified as ungraded, unless the 10th-grade enrollment was greater than zero.
- *Bureau of Indian Affairs (BIA) Schools.* These schools were identified using the state FIPS code = 59.
- *Special Education Schools.* Schools were classified as such if the indicator of school type was special education. Some schools for the blind and deaf were not indicated as special education, so schools with the words "blind," "unsighted," "deaf," or "impaired" in the school name were excluded (after manual verification).
- *Area Vocational Schools Not Enrolling Students Directly.* Public schools were classified as such if the indicator of school type was vocational and total enrollment was zero.
- *Schools That Are Detention Centers or Correctional Facilities.* Schools with the words "detention," "correctional," or "jail" in the school name were excluded (after manual verification).
- *Department of Defense (DOD) Schools Outside of the United States.* These schools were identified using the state FIPS code = 58.
- *Closed Public Schools.* These schools were identified using the status code on the CCD. Closed private schools were not able to be identified on the PSS.

If 10th-grade enrollment was unavailable on the school sample frame for any school, the enrollment for 10th grade was imputed based on the school's total enrollment, if known, or otherwise by using the median value of the enrollment for that grade for the appropriate stratum.

New high schools are not very common, and they are most common for small private schools. Schools were selected from a sampling frame that was 2 years old during the school year of the study, so a frame comparison was conducted between the 1998–99 CCD and the 1999–2000 CCD, and between the 1997–98 PSS and the 1999–2000 PSS, to determine the frequency of new high schools. Approximately 12 percent of the ELS-eligible public schools on the 1999–2000 CCD were not on the 1997–98 CCD, accounting for about 4 percent of the students. Approximately 21 percent of the ELS-eligible private schools on the 1999–2000 PSS were not on the 1997–98 PSS, accounting for about 8 percent of the students. It was therefore determined that it was necessary to update the sampling frame by adding some new schools.

To construct a supplemental sampling frame of new schools not currently on the sampling frame, a subsample of the public schools was selected. Each district associated with this new subsample of schools was considered in sample and was asked to identify public schools in their jurisdiction that taught 10th graders, that had recently opened, or that had begun

teaching 10th graders in the last 2 years. The districts were provided with a list of all public schools on the sampling frame in their district to help them identify the appropriate schools.

Similarly, a subsample of the Catholic sample schools was selected. Each diocese associated with this new sample of schools was considered in sample and was asked to identify Catholic schools in their jurisdiction that taught 10th graders, that had recently opened, or that had begun teaching 10th graders in the last 2 years. Each diocese was provided a list of all Catholic schools on the sampling frame in their diocese to help them identify the appropriate schools.

To identify other new private schools, the list frame for the 2001 PSS was used. There were approximately 1,400 new schools with 10th-grade students and approximately 900 new schools with an unknown grade span. Therefore, a larger sample of other new private schools was selected to account for the potentially high ineligibility rate. The new PSS schools were identified in spring 2001, so the list was about 1 year out of date at the time of ELS:2002 data collection. There was no perfect method to determine other new private schools for the 2001–02 school year, but public school districts were asked if they could identify new private schools in their area that enrolled 10th-grade students. The districts were provided with a list of the private schools in the district's area (based on zip code) that were either on the sampling frame or on the new PSS list. Some districts were able to identify the new private schools, but many districts were not able to do so. It was verified that all new schools identified to be on the supplemental sampling frame were not on the original sampling frame.

The field test sample was selected in such a way as not to adversely affect the full-scale sample. First, several schools with a large enrollment that could potentially be selected with certainty in the full-scale study were excluded. To determine these schools, a sample frame was formed that was similar to the one to be used in the full-scale study, each school's composite measure of size (MOS) was computed (see appendix J), and it was determined which schools would potentially be selected with certainty based on this MOS.

Second, the field test sample was designed such that schools selected for the field test would not be in the full-scale sample. For the field test, a stratified simple random sample of schools was selected using strata similar to those later used in the full-scale study. No probability-based inferences were made for the field test, even though a probability-based sample was selected, because the sample was too small to support such inferences. The objective was to have the complement of the field test sample, which was used for the full-scale study, to be a probability-based sample. The key fact that made this procedure work was that the complement of a simple random sample is also a simple random sample, and therefore is representative of the full population. For the full-scale study, field test sample schools were deleted from the frame, and each school on the sampling frame received a first-stage sampling weight based on the probability that it was not selected for the field test. An important benefit of this method of selecting the schools for the field test was that more recent versions of the CCD and PSS could be used for the full-scale sampling frame (e.g., the 1999–2000 CCD and PSS) without losing the ability to generalize to the full population. This method made no assumptions for the field test and full-scale study sampling frames. The impact of a school closing between the field test and full-scale study was negligible, since a PPS sample of schools was selected for the full-scale study. For the sample to be properly allocated for the full-scale study, the sample

was allocated before deleting the field test sample schools from the frame, and the full-scale strata included the field test strata. The NCES unique school identification numbers were used when matching the field test frame to the full-scale frame. Nonmatches within a state were sorted by school name and other fields as necessary and then manually checked for additional matches.

3.2.2 Stratification and School Selection

A stratified PPS sample of schools was selected using a composite size measure methodology developed by RTI statisticians (see appendix J).[19] A sample of approximately 800 (600 public, 200 private) schools from the school sampling frame was selected. The sampling frame for public schools was stratified by the nine-level U.S. Census divisions defined as follows:

- New England/Middle Atlantic—CT, ME, MA, NH, NJ, NY, PA, RI, VT;
- East North Central—IL, IN, MI, OH, WI;
- West North Central—IA, KS, MN, MO, NE, ND, SD;
- South Atlantic—DE, DC, FL, GA, MD, NC, SC, VA, WV;
- East South Central—AL, KY, MS, TN;
- West South Central—AR, LA, OK, TX;
- Mountain—AZ, CO, ID, MT, NV, NM, UT, WY; and
- Pacific—AK, CA, HI, OR, WA.

The New England and Middle Atlantic Census divisions were combined to be consistent with the NELS:88 stratification. Each geocode that contains a field test state (FL, IL, NC, NY, and TX) was substratified so that the school sample was correctly allocated and selected. States that were expected to have a public school sample size of at least 30 and therefore to have a state-representative sample were also substratified. Based on the 1997–98 CCD, CA, FL, NY, and TX were expected to have state-representative samples. Three of these states already were substrata because they were in the field test. The substrata were each state in the field test, each state with a state-representative sample, and all other states. For example, the South Atlantic was substratified by NC, FL, and all other states. Within each of these public school divisional strata or substrata, stratifications were made by metropolitan status based on CCD locale codes and defined as follows:

- *Urban:* the school is in a large or mid-size central city;
- *Suburban:* the school is in a large or small town or is on the urban fringe of a large or midsize city; and
- *Rural:* the school is in a rural area, either inside or outside a metropolitan statistical area (MSA).

[19] See Folsom, Potter, and Williams (1987).

These definitions are consistent with the NELS:88 stratification. Within each explicit stratum, implicit stratifications were made by state.

The sampling frame for Catholic and other private schools was stratified by Catholic and other private schools. Catholic schools were identified as those schools with affiliation identified on the PSS as Roman Catholic. Stratifications were then made by the four-level Census regions, defined as follows:

- Northeast—CT, ME, MA, NH, NJ, NY, PA, RI, VT;
- Midwest—IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI;
- South—AL, AR, DE, DC, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV; and
- West—AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY.

Each region that contains a field test state was substratified, so that the school sample was correctly allocated and selected. There were no states with a private school state-representative sample. The substrata were each state in the field test and all other states. For example, the South was substratified by NC, FL, TX, and all other states. Within each of these private school regional strata or substrata, the private schools were stratified by metropolitan status based on PSS data and defined similarly to the public school metropolitan status strata. Within each explicit stratum, implicit stratifications were made by religious affiliation.

Six hundred participating public schools were allocated proportional to the number of 10th-grade students contained within each public school stratum or substratum. One hundred participating Catholic schools and 100 participating other private schools were allocated proportional to the number of 10th-grade students contained within each Catholic and other private school stratum or substratum, respectively.

A sample size larger than 800 schools was necessary to compensate for the anticipated nonresponse. 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. To determine the overall sample size, assumptions were made about the expected response rate. Based on historical response rates in NELS:88 and HS&B, an overall response rate of approximately 70 percent was expected, but a response rate greater than 70 percent was attempted. It was planned to increase the sample size to 1,600, since there was uncertainty about achieving a response rate of at least 70 percent (sample size of 1,143). Such a response rate turned out to be harder to achieve than was the case in NELS:88 and HS&B (see table 43 in chapter 4). Table 11 gives the counts and percentages of sampled, eligible, and participating schools. These results show the difficulty of getting different types of schools to participate.

Table 11. School sampling, eligibility, and participation, by sampling stratum: 2002

School sampling stratum	Sampled schools		Eligible schools		Participating schools	
	Number	Percent[1]	Number	Percent[2]	Number	Percent[3]
Total	1,268		1221	96.29	752	61.59
Public	953	75.16	926	97.17	580	62.63
Catholic	140	11.04	140	100.00	95	67.86
Other private	175	13.80	155	88.57	77	49.68
Urban	434	34.23	414	95.39	250	60.39
Suburban	630	49.68	609	96.67	361	59.28
Rural	204	16.09	198	97.06	141	71.21

[1] Percent is based on overall total within column. Details may not sum to 100 percent due to rounding.
[2] Percent is based on number sampled within row.
[3] Percent is based on number eligible within row.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

A probability proportional to size (PPS) sample of all 1,600 schools was selected, as described above. One thousand of the 1,600 schools were randomly selected for the first release pool. Based on response rates from NELS:88, rates possibly higher than 80 percent were expected in a few strata (such as urban Catholic schools in the Northeast) and rates much lower than 80 percent were expected in a few strata (such as other private schools in the Northeast). Therefore, the 1,000 schools in the first release pool assumed an 80 percent overall response rate, but the sample size in some strata assumed greater than an 80 percent response rate. The sample size in some other strata assumed less than an 80 percent response rate. One hundred and forty-three schools were randomly selected from the remaining 600 schools for the second release pool to get a 70 percent overall response rate, and all remaining 457 schools were in the reserve pool.

Special efforts were made to minimize school-level overlap between ELS:2002 and the National Assessment of Educational Progress (NAEP). NAEP has both a national sample (at grades 4, 8, and 12), and numerous state samples (at grades 4 and 8). The ELS:2002 school sample was selected before the NAEP 2002 national school sample was selected. When the NAEP sample was selected, ELS:2002 selections were taken into account so that overlap between the two samples could be minimized.[20] Overlap with schools in the 2002 state NAEP sample was not minimized, since the state NAEP sample students were usually not in schools that included 10th grade. For ELS:2002 schools that did overlap with schools in state NAEP, the NAEP students and most teachers, given the different grade foci of the two studies, were not involved in ELS:2002. Table 12 summarizes the overlap with NAEP.

[20] When the ELS:2002 and NAEP probabilities of selection summed to less than 1.0, the NAEP sample excluded the ELS:2002 sample school. The NAEP probabilities of schools not selected in ELS:2002 were increased by 1/(1-P), where P was the ELS:2002 selection probability. When the ELS:2002 and NAEP probabilities of selection summed to 1.0 or greater, the NAEP probability of selection was reduced, but the ELS sample school was not necessarily excluded from the NAEP sample.

Table 12. School overlap between the ELS:2002 and the 2002 NAEP: 2002

School type	Total	NAEP sample[1]	
		Grade 4 or 8	Grade 12
Total	50	47	3
Public	29	29	0
Catholic/Other private	21	18	3

[1] The grade 4 and grade 8 NAEP samples are the state samples, and the grade 12 sample is the national sample.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Sample schools were later recruited for their participation in the study. Additionally, the associated school districts and dioceses for the public and Catholic sample schools, respectively, were recruited for their cooperation to allow the sample schools to participate, but no district-level interviews were attempted. Therefore, after the sample of public schools was selected, it was determined in which school districts the public sample schools were located and in which dioceses the Catholic schools were located. The sampling frame did not contain all contact information necessary for the initial contacts with schools, districts, and dioceses. However, the QED file contained contact information, so a database of schools, school districts, and dioceses was purchased from QED. The sample schools were matched with the QED database on CCD or PSS school ID and other fields, as necessary, such as name, address, and phone number. For matching schools, the principal's name was obtained from the QED file. Associated districts and dioceses were matched with the QED database on CCD or PSS district ID and other fields, as necessary, such as name, address, and phone number. For matching public districts, the superintendent's name was obtained from the QED file, and for matching dioceses, the contact's name, phone number, and address were obtained from the QED file, since these were not available on the PSS file. For schools, public districts, and dioceses that did not match to the QED, current principal, superintendent, or contact information was obtained from the Internet or from the district, diocese, and/or school through initial telephone contacts made by RTI recruiters. For example, the American Schools Directory (ASD) on the Internet was used to obtain principal names for public and private schools.

As described in section 3.2.1, the initial design was a sample of schools, so the probability of selecting a district (diocese) became a function of the number of schools in the district (diocese) and was incidental to the process of school sample selection. In addition, sometimes more than one school was selected from a district (diocese). While the initial sample selection was based directly on a sample of schools selected from a list of schools without regard to district (diocese) affiliation, the supplement of new schools depended on information provided at the district (diocese) level. The selection of a supplemental sample of new schools attempted to achieve two goals:

- to achieve unbiased estimation with school weights at about the level that would have resulted had they been included on the frame originally; and
- to determine the appropriate new school probability of selection from available data about the initial sampling frames and the new 10th-grade schools identified in the district (diocese).

To set a target selection probability for each new 10th-grade school identified by a district (diocese), the following were obtained or imputed:

- the primary school stratum, r ; and
- the 10th-grade enrollment for each race with a separate target sampling rate.

The target selection probability for a new school was then determined by substituting the new composite measure of size for school i in stratum r, $S_r(i)$, in the formula for probability of selection on school i in stratum r, $\pi_r(i)$ (see appendix J,) keeping all other factors (including the sum in the denominator) unchanged and calling the target probability of selection $\pi_{\text{r target (i)}}$.

For selecting the sample of new schools, a simple Bernoulli trial method of selecting the sample of supplemental schools was used.[21] Permanent random numbers were also assigned to all new schools associated with a district (diocese) and the sample was adjusted by adjusting the target selection probabilities as the number of schools selected within the district (diocese) increased.

3.2.3 Selection of Students

3.2.3.1 Sample Sizes

A sample of approximately 26 sophomores, from within each of the participating public and private schools was selected. Each school was asked to provide a list of 10th-grade students, and quality assurance (QA) checks were performed on each list that was received. A stratified systematic sample of students was selected on a flow basis as student lists were received. The strata were Hispanic, Asian, Black, and Other race/ethnicity.

The total expected student sample size of approximately 20,000 (approximately 800 x 25) was expanded to select additional Hispanic (if necessary) and Asian students to estimate subpopulation parameters within precision requirements. Table 13 lists these precision requirements, along with required sample sizes to meet the requirements. The required sample sizes were calculated under the following assumptions:

- use of two-tailed tests with significance of alpha = 0.05 to test differences between means and proportions with required power of 80 percent;
- use of a value of p = 0.30 to calculate sample sizes for estimates and tests of proportions;
- use of a mean value of 50 with standard deviation of 15 to calculate sample size for estimates and tests of mean;
- design effect is 2.0; and
- correlation between the main study and the first follow-up study is 0.6.

[21] Sometimes called Poisson sampling.

Table 13. Domain sample size requirements: 2002

Precision requirement	Required respondent sample size
Detect a 15 percent change in proportions across waves with 80 percent power using a two-tailed alpha = 0.05 test	1,356
Detect a 5 percent change in means across waves with 80 percent power using a two-tailed alpha = 0.05 test	454
Produce relative standard errors of 10 percent or less for proportion estimates based on a single wave of data	467
Produce relative standard errors of 2.5 percent or less for mean estimates based on a single wave of data	288

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

The largest required sample size (N = 1,356 respondents at the end of all follow-ups) was chosen for subpopulation estimation. Oversampling was used to try to ensure that each of the subpopulations had a minimum sample size of 1,356. Hence, it was attempted to achieve the precision requirements as follows:

- detect a 15 percent change in proportions across waves with 80 percent power using a two-tailed alpha = 0.05 test
- detect a 5 percent change in means across waves with 99 percent power using a two-tailed alpha = 0.05 test;
- produce relative standard errors of 6 percent or less for proportion estimates based on a single wave of data; and
- produce relative standard errors of 1.25 percent or less for mean estimates based on a single wave of data.

This sample size was inflated by the expected base year eligibility rates and by student response rates at the baseline and over four possible follow-up studies. This gave an initial (baseline) sample size of 2,193 Asian, 2,257 Hispanic, and 2,199 Black students, as shown in table 14. The base year eligibility and response turned out to be unrealistic, which was partially responsible for the student domain yields being lower than expected (see section 3.2.3.4 for additional discussion of student yield). Approximations using the race/ethnic percentages for public schools from the 1999–2000 CCD indicated that in a sample of approximately 15,000 public school students (approximately 600 x 25), it was expected to sample a minimum of 651 Asian students, 2,042 Hispanic students, 2,380 Black students, and 9,927 Others, without any oversampling. (The file indicated that about 4.3 percent of public school students are Asian, 13.6 percent are Hispanic, and 15.9 percent are Black.) Thus, we increased the sample size to include additional public school students in the sample. The total initial expected student sample size was approximately 21,757 (approximately 20,000 + [2,193 – 651] + [2,257 – 2,042]). A sample size of 2,193 Asians and 2,257 Hispanics was allocated so that the precision requirements could be met. The remaining sample size was allocated proportionally to the Black and Other race student strata. After the selection of student samples was begun, the sample rates were

adjusted, when necessary, to increase the actual number of expected Asians, Hispanics, and Blacks in the sample schools.

Table 14. Projected sample sizes, by race/ethnicity domains: 2002

	Asian	Black	Hispanic
Initial sample size	2,193	2,199	2,257
Base year eligibility rate[1]	93.60	94.30	96.40
Eligibles in base year	2,053	2,074	2,176
Base year response rate[1]	90.12	93.63	90.24
Respondents in base year	1,850	1,942	1,964
First follow-up response rate[1]	92.96	93.89	92.75
Respondents in first follow-up	1,720	1,823	1,822
Second follow-up response rate[1]	92.70	90.50	89.80
Respondents in second follow-up	1,594	1,650	1,636
Third follow-up response rate[1]	94.53	91.36	92.09
Respondents in third follow-up	1,507	1,507	1,507
Fourth follow-up response rate[2]	90.00	90.00	90.00
Respondents in fourth follow-up	1,356	1,356	1,356

[1]Expected rates at the time of sampling based on National Education Longitudinal Study of 1988 (NELS:88) unweighted response rates. Assumed response rate for Asian-Pacific Islanders for Asian response rate. Assumed response rate for Others in base year for Black response rate.
[2]Response rates for fourth follow-up are minimum expected rates based on NELS:88 first three follow-ups.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.2.3.2 Specifications for Enrollment Lists

Each sample school was asked to provide an electronic or hardcopy listing of all their 10th-grade students currently enrolled.

The information requested for each eligible student was as follows:

- student ID number;
- Social Security number (may be the same as the ID number; this item was optional);
- full name;
- sex;
- race/ethnicity (White; Black; Asian; Native Hawaiian or Other Pacific Islander; American Indian or Alaska Native; Hispanic; Other); and
- whether or not an individualized education program (IEP) was currently filed for the student (yes, no).

The race/ethnicity variables were needed to allow for oversampling of Asians and Hispanics. The race, ethnicity, sex, and IEP variables were potentially useful for nonresponse adjustments.

It was requested that the electronic file be a column-formatted or comma-delimited American Standard Code for Information Interchange (ASCII) file or Excel file. Schools provided the electronic lists via e-mail, provided a diskette containing the file, or uploaded the

file to the ELS:2002 web site. If the school could not provide an electronic file, it was asked to provide a hardcopy list, preferably in alphabetical order within race/ethnicity strata to facilitate stratified systematic sampling. Whatever the school could provide was accepted to select the student samples; however, every effort was made to facilitate receiving uniformly formatted electronic files from as many schools as possible to make processing them quicker, more reliable, and less expensive.

The specifications for the list request were presented and their importance explained in the school coordinator's packet. In addition to the items described above, the coordinator's packet contained detailed instructions for preparing the student lists, school ID labels to place on all diskettes and hardcopy lists, an express mail airbill, and instructions for sending the file layouts and data files to RTI via e-mail or uploading if any of those options was desired. The detailed instructions included guidelines identifying the eligible students and data elements to be listed by the school, completed hardcopy examples, a transmittal form with the file layout for electronic files, and a checklist for proper completion of hardcopy lists. Table 15 shows the number and percentage of schools that provided lists and participated, and the types of student lists provided by schools.

Table 15. Types of student lists provided, by schools: 2002

Type of list received	Frequency[1]	Percent
Total	767	100.00
Both electronic and hardcopy	18	2.35
Electronic copy	378	49.28
Hardcopy	371	48.37

[1]The counts include all schools that sent in a list, but 15 of these schools later decided not to participate in ELS:2002.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.2.3.3 Quality Assurance Checks

Quality assurance (QA) checks were performed on all received lists. Any lists that were unreadable immediately failed the QA checks. Since the students were stratified by Hispanics, Asians, Blacks, and Other race/ethnicity, the list failed the QA checks if it did not allow stratification of the students.

The school's count of 10th-grade students was also checked to verify that the school provided a complete list of eligible students. For public and private schools, the provided counts of 10th-graders were compared with the total counts and counts by strata on the frame (CCD and PSS). The PSS does not provide counts by the strata, so the race/ethnicity breakdowns were estimated by assuming the percentage of students in the school of a certain race/ethnicity was similar to the percentage of that race/ethnicity for 10th graders. The CCD and PSS contain flags that identify whether the enrollment has been imputed. For schools with an imputed enrollment, the counts were not compared, and the list passed the QA check. If any of the counts of 10th graders for total students or by the race/ethnicity strata on the provided list were 25 percent lower or 25 percent higher than the frame counts, then the list failed the QA check, unless the provided count was greater than zero and the absolute difference was less than 100. However, if the

school count of Hispanics or Asians was zero and the frame count was less than five, the count did not fail the QA check.

Schools that failed the QA check were recontacted by the school recruiter to resolve the discrepancy and to verify that the school representative who prepared the student lists clearly understood the data collection request and provided lists of the eligible students. When it was determined that the initial list provided by the school was not satisfactory, a replacement list was requested. If the school confirmed that the list was correct or if the school sent a replacement list, selection of the sample students proceeded. If the school refused to send a replacement list, then selection of the sample students also proceeded. Table 16 lists the frequency of types of problems encountered with student lists.

Table 16. Types of problems encountered with student lists: 2002

Type of problem	Frequency	Percent
Total	752	100.00
None	530	70.48
Unreadable file or list	13	1.73
Count out of bounds	40	5.32
Cannot identify strata	142	18.88
Insufficient documentation	5	0.66
Multiple problems	13	1.73
Other problem	9	1.20

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.2.3.4 Student Sampling from Lists

Students from schools were sampled on a flow basis as student lists were received. Stratified systematic sampling procedures were used for both electronic and hardcopy lists. For each school, the student sample rates were fixed, rather than the student sample sizes, for the following reasons:

- to facilitate sampling students on a flow basis as we received student lists; and
- because sampling at a fixed rate based on the overall student stratum sampling rate and the school probabilities of selection would result in approximately equal overall probabilities of selection within the ultimate school by student strata. (See appendix J for mathematical details of student sampling.)

Each time schools were released from the second release pool or the reserve sample pool, sampling rates were adjusted to account for the nonresponding schools and the new schools.

For schools that provided electronic lists of students, the lists were stratified by race/ethnicity within grade level and a stratified systematic sample of students was selected.

For schools that provided hardcopy lists, an efficient two-stage process was used to select systematic samples from hardcopy lists. Sample pages were first selected and then sample students within strata were chosen from the selected pages. The page sampling rate was set so

that approximately 10 students were selected from each page. This approach was particularly efficient for long lists. The sample was keyed after it was selected.

When a hardcopy list included Hispanic and Other race students together who had to be sampled at different rates, the list was initially sampled at the higher student sampling rate. Then, the initial sample was keyed, the stratum which had the lower sampling rates was subsampled to achieve the proper sample inclusion rates. When a hardcopy list included Asian students not separated from the other students, a student identifier was keyed for these Asian students and a systematic sample was separately selected. This helped avoid potential sample size and precision problems for the Asian students that might have occurred due to clustering of last names on the enrollment list.

After the student sample was selected, it was verified that the sample size was within reasonable bounds of the school's expected sample size. If the total number of sample students was fewer than 10 (unless all students had been selected), or if the number selected was greater than 35, the sampling rates were adjusted accordingly and the sample was reselected. Table 17 shows the numbers of students sampled and eligible sample students. The sample counts are generally less than the expected counts for four main reasons. First, students were sampled from 752 schools rather than from 800 schools as planned (see table 43 in chapter 4). Second, the planned sampling rates frequently would have given a sample greater than the maximum size of 35, so the sampling rates were often trimmed to achieve a size of 35. Third, the ineligibility rate was higher than expected. Fourth, the expected numbers of certain student population domains at some schools were lower than expected. Adjustments were made to sampling rates to schools later in the process in an attempt to help account for the lower domain and overall sample sizes.

3.2.3.5 Sample Updating

The student sample was selected, when possible, in the fall or early winter so that sample teachers could be identified (see section 3.2.5) and materials could be prepared well in advance of Survey Day. However, selecting the sample in advance meant that some students transferred into the sample schools and others left between the time of sample selection and Survey Day.

In previous studies such as HS&B and NELS:88, as part of the sample updating procedure, schools were asked to supply a list of students in the indicated grade who had newly enrolled in the school since the time that the original sample had been drawn. Analysis of such lists both in NELS:88[22] and in the NAEP trial assessments[23] suggested that there was systematic and serious underreporting of students who had transferred in. To address this problem, complete enrollment lists were collected at both the time of initial sampling and the time of the sample update.

For identifying students who transferred into the school since the first list was prepared, a technique known as the "half-open interval rule" was used. The steps were similar to those for "freshening" the sample with 12th graders in the first follow-up. At the time of the initial request for the student lists, the school was informed that a second list of students would be necessary approximately 3 weeks prior to data collection to allow sample updating. If the school required

[22] See Ingels, Scott, and Taylor (1998).
[23] See Spencer (1991).

Table 17. Expected and achieved student samples, by student stratum: 2002

Student type	Number expected					Number achieved					Number eligible				
	Total	Hispanic	Asian	Black	Other	Total	Hispanic	Asian	Black	Other	Total	Hispanic	Asian	Black	Other
Total	21,759	2,646	2,441	2,750	13,922	19,218	2,250	2,014	2,657	12,297	17,591	2,001	1,891	2,323	11,376
Public	16,758	2,257	2,193	2,380	9,928	15,361	2,020	1,860	2,382	9,099	13,882	1,780	1,744	2,070	8,288
Catholic	2,501	268	119	187	1,927	2,156	191	83	165	1,718	2,113	187	78	159	1,689
Other private	2,500	121	129	183	2,067	1,701	39	72	110	1,480	1,596	34	69	94	1,399

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

explicit parental consent, then the second list was requested approximately 5 weeks prior to data collection in order to allow enough time to resolve issues related to obtaining permission for students to be in the study. This second list allowed transfer students the opportunity to be selected. The steps in the procedure were as follows:

Step 1: The recruiter requested an updated list of all 10th-grade students. If the school provided electronic lists, then both the first and second lists were sorted in the same order. If the school sent hardcopy lists for both the first and second lists, then the school needed to sort the second list in the same way as the first list (e.g., both sorted alphabetically for each stratum).

Step 2: Quality assurance (QA) checks and problem resolution were performed in a manner similar to the procedures for the original lists described previously in this chapter. The counts of students within each stratum were expected to be similar to the counts on the first list. If any of the counts of 10th graders for total students or by the race/ethnicity strata on the updated list were 25 percent lower or 25 percent higher than the counts on the original list, then the list failed the QA check unless the provided count was greater than zero and the absolute difference was less than 50. However, if the updated count of Hispanics or Asians was zero, and the original count was less than three, the count did not fail the QA check.

Step 3: The sampled ELS:2002 students were identified on the new list. For students not on this list, it was determined where they would have been on the list if they were still enrolled.

Step 4: To select transfer students at the same rate as the initial sample, the first requested student lists from which the sample of approximately 25 10th graders were selected were compared to the second lists. If the person immediately following each sampled individual within the race/ethnicity strata[24] on the second list was not on the first list (for whatever reason), it was assumed that the student was a transfer student, and that student was included in the sample. If the last student on the list was a sampled student, then the next student was the first student on the list (i.e., the list was "circularized").

Step 5: Whenever a transfer student was added to the sample, it was determined whether the next student on the roster was a transfer student or not. Once a student who was not a transfer student had been identified, then the process continued for the next sample student on the roster. The sequence of steps 4 and 5 was continued, and more transfer students were added, until a student who was enrolled at the time of the initial list was reached on the roster.

These second lists were also used to identify students who were no longer at the school. If a sample student was not on the second list, then that student was no longer at the school and

[24] Race/ethnicity strata for students on both the original and new lists were based on the original list used for sampling, even if the student's race/ethnicity was reported differently on the new list.

no longer in the sample. However, the check for transfer students was still implemented on the basis of where the student would have been on the second list, if the student was still enrolled.

Not as many updated lists were received as anticipated for two reasons. First, it was expected that most schools would send original enrollment lists in the fall and updated enrollment lists in the spring. However, many schools sent original lists in winter or spring close to the time of Survey Day, so there was no time for them to send an updated list. Second, at the time updated lists were requested, many schools were preparing lists of teachers and addresses of sample students and were too busy to send an updated list. From the 123 updated lists received, 86 students were added (0.70 per school).

3.2.4 Student Eligibility and Exclusion

All spring-term 2002 sophomores in eligible schools, except for foreign exchange students, were eligible for the study. This meant that several categories of students who were ineligible for HS&B and NELS:88 were eligible for ELS:2002 (though it did not mean that such students were necessarily tested or that they completed questionnaires).

In NELS:88, the following categories of students were deemed ineligible:

- students with disabilities (including students with physical or mental disabilities, or serious emotional disturbance, and who normally had an assigned IEP) whose degree of disability was deemed by school officials to make it impractical or inadvisable to assess them; and
- students whose command of the English language was insufficient, in the judgment of school officials, for understanding the survey materials, and who therefore could not validly be assessed in English.

In ELS:2002, the treatment of these categories of students was addressed as discussed below.

3.2.4.1 Schools Given Clear Criteria for Including/Excluding Students

Students were not excluded categorically (e.g., just because they received special education services, had IEPs, received bilingual education or English as a second language services), but rather on a case-by-case (individual) basis. The guiding assumption was that many students with IEPs or limited English proficiency (LEP) would be able to participate, and schools were requested if unsure, to include the student. Although both questionnaire and assessment data were sought, the minimum case of participation was completion of the student questionnaire. Hence some students who could not be assessed could nevertheless participate; that is, complete the questionnaire.

In addition, the ELS:2002 assessments were more accessible to many students who formerly (as in NELS:88) might have been excluded, for two reasons in particular. First, the ELS:2002 base year test was two-stage and adaptive, unlike the base year NELS:88 test; second, unlike NELS:88, ELS:2002 offered various testing accommodations.

The ELS:2002 test battery was an adaptive test, and thus better suited to students with learning disabilities than would be a conventional test. The ELS:2002 battery was a two-stage assessment (routing test and second-stage test tailored to the ability level determined in the routing test). Because it was designed to avoid floor effects, it contained many items that were well below grade level. Because the test was adaptive, it could route students of lower achievement to a simpler second-stage form of the test (i.e., one with easier items that better corresponded to their mastery level).

Several testing accommodations were also provided. Schools and parents were urged to permit the study to survey and test students under these special conditions.

The suggested criterion for exclusion of students from survey instrument completion on language grounds followed the current practice for the NAEP students. Students were regarded as capable of taking part in the survey session (test and questionnaire administration) if they had received academic instruction primarily in English for at least 3 years or they had received academic instruction in English for less than 3 years, but school staff judged or determined that they were capable of participating. In terms of exclusion from taking the instruments on disability grounds, it was suggested that only if the student's IEP specifically recommended against their participation in assessment programs should they be excluded, and then only from the tests, if questionnaire level participation were possible. Moreover, if their IEP stated that they could be assessed if accommodations were provided, then their participation became a question of whether the school could supply the particular accommodation. The specific accommodations offered by schools are set out immediately below, under the second point of this discussion.

3.2.4.2 Accommodations Offered to Increase Participation

To the extent possible, given practical and monetary constraints, accommodations were offered to increase the number of participants. All tests taken under conditions of special accommodations were flagged on the data file (BYTXACC is the accommodation indicator), and the nature of the accommodation was noted.

In theory, many kinds of accommodations were possible. There were accommodations of test presentation, of response, of setting, and of allotted testing time. In addition to accommodations for the assessments, special measures were employed to facilitate questionnaire completion (e.g., in some instances, ELS:2002 students were administered the student questionnaire by survey staff, if self-administration was not possible for them).

One type of accommodation offered is alternative test presentation (e.g., on mathematics tests, one might read problems aloud, have someone sign the directions using American Sign Language, use a taped version of the test, provide a Braille or large-print edition of the test, or supply magnifying equipment). While the study could not, for example, provide Braille translations, when a school could assist in providing a presentational accommodation (as with magnifying equipment or an aide who translated directions into American Sign Language), its use was deemed an acceptable accommodation.

A second type of accommodation sometimes offered is alternative means of test responses (e.g., responses made in Braille or American Sign Language or produced using a keyboard or specially designed writing tool). However, ELS:2002 was not able to provide special accommodations for responding.

A third type of accommodation sometimes offered is providing an alternative setting. For example, an emotionally disturbed student might not be a good candidate for a group administration, but might be assessed alone. ELS:2002 made this type of accommodation available where possible or permissible by the school.

A fourth possible kind of accommodation is in timing or length of administration (or length of any given test session). Although tests were strictly timed in the three prior high school longitudinal studies, giving extra time posed less of a threat to validity for ELS:2002, given that it was an adaptive test, and that extra time could be restricted to the second stage of the test. There were three options for proceeding—give extra time in one session; keep testing time constant in minutes tested but give more breaks, or split test sessions over several days. Table 18 lists the counts for students excluded from survey instrument completion and students accommodated.

3.2.4.3 Questionnaire Eligibility Status to Be Reassessed in the First Follow-up

A special substudy of excluded students was conducted in NELS:88.[25] It was found that there was considerable change in eligibility status, especially for students excluded for reasons of their English language proficiency, across rounds (e.g., 71 percent of base year excluded LEPs became eligible over time, as did 57 percent of the entire excluded group). Since for ELS:2002, like NELS:88, the sample design calls for generating representative senior cohorts as well as sophomore cohorts, these status changes should be taken into account. Moreover, the senior year will be treated as a baseline for a new panel (i.e., 2004 seniors), making data collected from excluded sophomores who progress to senior year in the modal sequence fully usable for longitudinal analysis of the senior cohort.

3.2.4.4 Enrollment Status, Records, and Contextual Data Gathered for Students Unable to Be Surveyed or Validly Assessed

In addition to documenting the reasons test-exempted students could not be assessed, their enrollment status will be tracked so that it is known whether they are in school or are dropouts 2 years later. Parent questionnaires and teacher reports were collected for these students in the base year. In the first follow-up, high school transcripts will be collected for these students as well. School-level data, such as school administrator survey responses in the base year and first follow-up, will also be linked to these students. A contextual or expanded sample weight—as contrasted to the student questionnaire completion weight—has been created and is included on the restricted-use data file. The expanded sample weight generalizes to all spring term 2002 sophomores and will facilitate analysis of students who were exempted from completing the survey forms.

[25] See Ingels (1996).

Table 18. Counts of students excluded and students accommodated: 2002

Excluded or accommodated	Count
Number of students excluded	163
Mental or physical disability	119
Language barrier (LEP/NEP)[1]	44
Number of students accommodated	114

[1]LEP=limited English proficient; NEP=non-English proficient.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.2.5 Selection of Contextual Samples

As described in section 2.4, ELS:2002 included a teacher survey that gathered teacher reports on students' learning experiences and performance. Teachers in two subject areas (mathematics and English) were eligible if they taught students who were sampled for ELS:2002.

Some sample students may have had multiple or zero mathematics or English teachers during the 2001–02 school year (e.g., different teachers for the fall and spring terms). In these situations, the fall-term teacher was used as the relevant reference point, if possible. It was decided as follows which mathematics or English teacher, if any, to include in the teacher sample:

- If fall teacher A and spring teacher B, then sampled fall teacher A;
- If fall teacher A left the school and spring teacher B was present, then sampled spring teacher B;
- If no fall teacher but one spring teacher, then sampled spring teacher;
- If no fall teacher but two or more spring teachers, then randomly selected one to be in sample;
- If no spring teacher but fall teacher, then sampled fall teacher;
- If two or more fall teachers, then randomly selected one to be in sample; and
- If no fall teacher and no spring teacher, then no teacher was in sample.

Table 19 shows the number of sample teachers who taught mathematics, English, or both subjects. The sample counts are also displayed by type of school and urbanicity.

Table 19. Sample teachers, by subject taught, school type, and school urbanicity: 2002

Teacher/school characteristic	Frequency	Percent	Average number per responding school
Total	9,287		12.62
Subject			
Math	5,090	54.8	6.92
English	4,027	43.5	5.49
Both	160	1.7	0.22
School type			
Public	8,237	88.7	14.55
Catholic	692	7.5	7.28
Other private	358	3.9	4.77
Urbanicity			
Urban	3,347	36.0	13.77
Suburban	4,480	48.2	12.69
Rural	1,460	15.7	10.43

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

For each sample student, there was one sample parent. The NELS:88 procedures were followed to identify the sample parent by asking which parent, in two-parent households, was most knowledgeable about the student's educational situation. For one-parent households, that parent was in the sample.

For each sample school, the principal and library media specialist were also in the sample.

3.2.6 Bias Analysis for School and Student Unit Nonresponse

Unit nonresponse causes bias in survey estimates when the outcomes of respondents and nonrespondents are different. ELS:2002, has two levels of unit response: 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 the 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 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. For students, although the overall unweighted 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. See section 4.9 for a further discussion of response rates.

Extensive data were available for nonresponding schools, which were used to help reduce potential nonresponse bias. Nonresponding schools (or their districts) were asked to complete a school characteristics questionnaire. (Of the 469 nonresponding eligible sample schools, a total of 437—or 93.18 percent—completed the special questionnaire).

The nonresponding school questionnaire contained a subset of questions from the school administrator questionnaire that was completed by the principals of participating schools. The school sampling frame constructed from the CCD and PSS also contained data for all schools. Usable data included the following:[26]

- school type
- metropolitan status
- region
- 10th-grade enrollment
- total enrollment
- number of minutes per class
- number of class periods
- number of school days
- number of students receiving free or reduced price lunch
- number of full-time teachers
- percentage of full-time teachers certified
- number of part-time teachers
- number of different grades taught at the school
- school level
- coeducational status
- percentage of students with an IEP
- percentage of students with LEP
- percentage Hispanic 10th-grade students
- percentage Asian 10th-grade students
- percentage Black 10th-grade students
- percentage White and Other race 10th-grade students.

Some information on 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). In consequence, 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.

First, for these data known for most respondents and nonrespondents, nonresponse bias was estimated and tested to determine if the bias was significant at the 5 percent level. Second, nonresponse adjustments were computed (see sections 3.4.1 and 3.4.2), and variables known for most respondents and nonrespondents (those listed above) were included in the nonresponse models (see section 3.4). The school and student nonresponse adjustments were designed to significantly reduce or eliminate nonresponse bias for variables included in the models. Variables not known for most respondents and nonrespondents could not be included in the nonresponse adjustments, and therefore nonresponse bias could not explicitly be reduced for these variables. However, the variables in the school nonresponse model are arguably the most analytically important school-level variables and are correlated with many of the other school-level variables. Likewise, many of the variables in the student nonresponse model are correlated with many of the other student-level variables.

[26] These variables were also used in the nonresponse weighting adjustment described in section 3.4.1.

Third, once the school and student weights (after nonresponse adjustment) were computed, remaining bias for data known for most respondents and nonrespondents was estimated and statistically tested to verify that there was no remaining significant nonresponse bias. Fourth, the remaining bias for all variables after student weight adjustments was divided by the standard error, that is, bias / standard error.

The nonresponse bias was estimated for variables known for both respondents and nonrespondents. The bias in an estimated mean based on respondents, $\bar{y}_R$, is the difference between this mean and the target parameter, π, i.e., the mean that would be estimated if a complete census of the target population was conducted. This bias can be expressed as follows:

$$B(\bar{y}_R) = \bar{y}_R - \pi .$$

The estimated mean based on nonrespondents, $\bar{y}_{NR}$, can be computed if data for the particular variable for most of the nonrespondents is available. The estimation of π is as follows:

$$\hat{\pi} = (1-\eta)\,\bar{y}_R + \eta\bar{y}_{NR}$$

where η is the weighted[27] unit nonresponse rate. For the variables that are from the frame rather than from the sample, π can be estimated without sampling error. The bias can then be estimated as follows:

$$\hat{B}(\bar{y}_R) = \bar{y}_R - \hat{\pi}$$

or equivalently

$$\hat{B}(\bar{y}_R) = \eta(\bar{y}_R - \bar{y}_{NR}) .$$

This formula shows that the estimate of the nonresponse bias is the difference between the mean for respondents and nonrespondents multiplied by the weighted nonresponse rate. The variance of the bias was computed using Taylor series estimation in RTI's SUDAAN software package.

Tables 20 and 21 show the nonresponse bias before and after weight adjustments (see section 3.4.1) for selected variables for all schools. The first set of columns shows the estimated bias before nonresponse adjustment for the variables available for most responding and nonresponding schools. Statistical tests (*t* tests) was used to test each level of the variables for significance of the bias at the 0.05/(c – 1) significance level, where c is the number of categories within the primary variable. Below is the summary of the before-adjustment significant bias at the school level:

- at least one level of most of the variables was biased,
- thirty-eight variables (continuous variables and levels of categorical variables) were found to be significantly biased, and
- significant biases were usually small.

[27] The weight used will be the weight prior to nonresponse adjustment, i.e., the school-level design weight multiplied by the first-stage sampling weight multiplied by the release weight (see section 3.4.1) for details of these weights).

The second set of columns in tables 20 and 21 shows the estimated bias after weight adjustments for the variables available for most responding and nonresponding schools. The bias after weight adjustments was computed as the difference between the estimate using nonresponse-adjusted (final) weights and the estimate using the design (base) weights prior to nonresponse adjustment. This latter estimate is an estimate of π because it is the estimate of the target population using the sample weights. Similar to the before-adjustment bias, *t* tests were performed to test the significance of the bias for each level of the variables, and Chi-square tests were performed to test the significance of the distributions of each variable. For school level nonresponse bias analysis, the estimated bias decreased after weight adjustments for many variables. In fact, the number of significantly biased variables decreased from 38 before adjustment to 1 after adjustment.

The one variable still showing significant bias after weight adjustments is the continuous variable 10th-grade enrollment. The nonresponse adjustment model could only use categorical variables as independent variables, and the 10th-grade enrollment categorical variable was included in the model and has no remaining significant bias.

Table 22 shows the nonresponse bias before and after weight adjustments (see section 3.4.2) for selected variables for all students. As is the case on the school level table, the first set of columns shows the estimated bias before nonresponse adjustment for the variables available for most responding and nonresponding students. Statistical tests (*t* tests) were used to test each level of the variables for significance of the bias at the $0.05/(c - 1)$ significance level, where c is the number of categories within the primary variable. Below is the summary of the before-adjustment significant bias for table 21 (student level):

- at least one level of most of the variables was biased;
- 42 variables were found to be significantly biased;
- significant biases were usually small.

As in tables 20 and 21, the second set of columns in table 22 shows the estimated bias after weight adjustments for the variables available for most responding and nonresponding students. The bias after weight adjustments was computed as the difference between the estimate using nonresponse-adjusted (final) weights and the estimate using the design (base) weights prior to nonresponse adjustment. This latter estimate is an estimate of π because it is the estimate of the target population using the design weights. Similar to the testing of before-adjustment bias, *t* tests were performed to test the significance of the bias for each level of the variables, and Chi-square tests were performed to test the significance of the distributions of each variable. For student level nonresponse bias analysis, the estimated bias decreased after weight adjustments for every variable. Therefore, the number of significantly biased variables decreased from 42 *before* adjustment to zero *after* adjustment.

Table 20. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for schools: 2002

Description / Response	Before nonresponse adjustment						After nonresponse adjustment			
	Unweighted respondents	Unweighted non-respondents	Respondent mean weighted[1]	Non-respondent mean weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Relative bias
Minutes per class period										
≤ 45	174	97	23.16	23.29	-0.05	-0.00	23.20	23.43	-0.23	-0.01
46–50	161	108	21.99	25.76	-1.31	-0.06	23.30	23.18	0.12	0.01
51–80	196	133	25.83	30.26	-1.54	-0.06	27.37	27.49	-0.12	-0.00
81+	207	86	29.03	20.7	2.90*	0.11	26.13	25.90	0.23	0.01
Class periods per day										
1–4	210	89	29.45	21.32	2.85*	0.11	26.60	26.61	-0.01	0.00
5–6	181	151	23.55	33.72	-3.56*	-0.13	27.11	27.27	-0.16	-0.01
7	203	110	27.89	26.45	0.50	0.02	27.39	27.58	-0.19	-0.01
8–9	142	77	19.1	18.51	0.21	0.01	18.90	18.54	0.36	0.02
Total enrollment										
≤ 600	185	96	25.07	21.02	1.48	0.06	23.59	23.59	0.00	0.00
601–1,200	219	121	29.8	27.09	0.99	0.03	28.81	28.90	-0.09	-0.00
1,201–1,800	168	100	22.81	22.17	0.23	0.01	22.57	22.54	0.03	0.00
> 1,800	172	144	22.33	29.71	-2.70*	-0.11	25.03	24.97	0.06	0.00
10th-grade enrollment										
0–99	160	69	21.38	16.74	1.60	0.08	19.79	20.34	-0.55	-0.03
100–249	187	93	25.36	22.33	1.04	0.04	24.32	24.12	0.20	0.01
250–499	240	133	32.00	31.53	0.16	0.01	31.84	31.59	0.25	0.01
500+	165	130	21.26	29.40	-2.80*	-0.12	24.06	23.95	0.11	0.01
Free or reduced-price lunch										
0	137	103	20.38	26.75	-2.23*	-0.10	22.61	23.40	-0.79	-0.04
1–10	150	93	21.54	23.57	-0.71	-0.03	22.25	21.79	0.46	0.02
11–30	196	110	28.85	28.13	0.25	0.01	28.60	28.11	0.49	0.02
> 30	202	89	29.23	21.55	2.69*	0.10	26.55	26.70	-0.15	-0.01
Number of full-time teachers										
1–40	195	78	27.42	18.89	3.00*	0.12	24.43	24.55	-0.12	-0.01
41–70	183	102	25.83	25.02	0.28	0.01	25.55	25.54	0.01	0.00
71–100	171	115	23.69	27.24	-1.25	-0.05	24.94	24.35	0.59	0.02
101+	166	123	23.05	28.85	-2.04*	-0.08	25.09	25.55	-0.46	-0.02
Number of grades within the school										
4	554	322	74.49	73.41	0.38	0.01	74.11	73.47	0.64	0.01
> or < 4	188	112	25.51	26.59	-0.38	-0.01	25.89	26.53	-0.64	-0.03
IEP[3] percentage										
≤ 5	281	148	40.38	37.16	1.13	0.03	39.25	39.60	-0.35	-0.01
6–10	176	137	25.25	32.96	-2.71*	-0.10	27.96	28.15	-0.19	-0.01
11–15	145	85	20.75	21.04	-0.10	-0.00	20.85	20.20	0.65	0.03
> 15	95	37	13.62	8.84	1.68*	0.14	11.94	12.05	-0.11	-0.01

See notes at end of table.

Table 20. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for schools: 2002–Continued

Description / Response	Before nonresponse adjustment						After nonresponse adjustment			
	Unweighted respondents	Unweighted non-respondents	Respondent mean weighted[1]	Non-respondent mean weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Relative bias
LEP[4] percentage										
0	327	152	46.57	38.17	2.93*	0.07	43.64	44.46	-0.82	-0.02
1	135	70	19.13	17.85	0.45	0.02	18.68	18.26	0.42	0.02
2–5	118	70	16.66	17.41	-0.26	-0.02	16.92	16.54	0.38	0.02
> 6	133	119	17.64	26.57	-3.11*	-0.15	20.75	20.74	0.01	0.00
Number of part-time teachers										
0–1	201	105	29.94	27.31	0.91	0.03	29.03	29.03	0.00	0.00
2–3	196	92	28.79	23.63	1.79	0.07	27.00	27.07	-0.07	-0.00
4–6	161	93	23.08	23.96	-0.30	-0.01	23.39	23.08	0.31	0.01
7+	127	102	18.18	25.11	-2.40*	-0.12	20.58	20.82	-0.24	-0.01
Full-time teachers certified										
0–90	182	109	25.45	25.91	-0.16	-0.01	25.62	25.61	0.01	0.00
91–99	125	81	17.83	19.21	-0.48	-0.03	18.31	18.15	0.16	0.01
100	401	222	56.72	54.88	0.65	0.01	56.07	56.24	-0.17	-0.00
Number of days in school year										
Less than 180	187	115	25.98	28.17	-0.76	-0.03	26.74	27.11	-0.37	-0.01
180	413	244	56.15	56.65	-0.17	-0.00	56.33	56.32	0.01	0.00
More than 180	135	65	17.87	15.18	0.94	0.06	16.93	16.57	0.36	0.02
Is the school coeducational?										
Yes	699	411	94.16	93.71	0.16	0.00	94.00	93.80	0.20	0.00
No, all-female school	19	9	2.69	2.17	0.18	0.07	2.51	2.78	-0.27	-0.11
No, all-male school	22	16	3.16	4.11	-0.34	-0.10	3.49	3.42	0.07	0.02
Type of grades within the school										
K–12, PreK–10th, 1st–12th, PreK/1st–9th/12th and PreK–12	57	59	7.56	13.95	-2.25*	-0.23	9.81	10.47	-0.66	-0.07
Middle grades but no elementary	79	32	10.73	7.50	1.14	0.12	9.59	9.51	0.08	0.01
Only high school	606	343	81.71	78.55	1.11	0.01	80.60	80.01	0.59	0.01
School type										
Public	580	346	76.76	72.90	1.42	0.02	75.34	75.34	0.00	0.00
Catholic	95	45	13.41	10.56	1.05	0.08	12.36	12.36	0.00	0.00
Other private	77	78	9.83	16.55	-2.47*	-0.20	12.30	12.30	0.00	0.00
Metropolitan status										
Urban	250	164	34.22	35.95	-0.64	-0.02	34.86	34.86	0.00	0.00
Suburban	361	248	46.05	50.68	-1.70	-0.04	47.75	47.75	0.00	0.00
Rural	141	57	19.73	13.37	2.34*	0.13	17.39	17.39	0.00	0.00

See notes at end of table.

Table 20. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for schools: 2002–Continued

Description / Response	Before nonresponse adjustment						After nonresponse adjustment			
	Unweighted respondents	Unweighted non-respondents	Respondent mean weighted[1]	Non-respondent mean weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Relative bias
Geocode										
Census division (public schools)										
Public—New England/Middle Atlantic[5]	95	82	11.26	16.21	-1.82*	-0.14	13.08	13.20	-0.12	-0.01
Public—East North Central	90	46	12.61	10.90	0.63	0.05	11.98	11.42	0.56	0.05
Public—West North Central	48	13	6.84	3.19	1.34*	0.24	5.50	6.62	-1.12	-0.20
Public—South Atlantic	117	30	16.83	7.32	3.50*	0.26	13.33	14.15	-0.82	-0.06
Public—East South Central	41	9	5.78	2.18	1.32*	0.30	4.46	4.70	-0.24	-0.05
Public—West South Central	69	35	9.72	8.48	0.46	0.05	9.26	7.90	1.36	0.15
Public—Mountain	34	25	4.74	6.01	-0.47	-0.09	5.21	5.64	-0.43	-0.08
Public—Pacific	86	106	8.98	18.60	-3.54*	-0.28	12.51	11.72	0.79	0.06
Census region (private schools)										
Private—Northeast	39	41	5.19	8.52	-1.22*	-0.19	6.41	6.29	0.12	0.02
Private—Midwest	51	27	6.38	5.91	0.17	0.03	6.21	5.66	0.55	0.09
Private—South	54	31	7.99	7.46	0.19	0.03	7.79	8.09	-0.30	-0.04
Private—West	28	24	3.68	5.22	-0.57	-0.13	4.25	4.62	-0.37	-0.09
Asian 10th-grade enrollment										
≤ 2 percent	292	148	39.80	33.51	2.31*	0.06	37.49	37.49	0.00	0.00
> 2 percent	460	321	60.20	66.49	-2.31*	-0.04	62.51	62.51	0.00	0.00
Black 10th-grade enrollment										
≤ 4 percent	255	207	33.31	43.67	-3.81*	-0.10	37.12	37.12	0.00	0.00
> 4 percent	497	262	66.69	56.33	3.81*	0.06	62.88	62.88	0.00	0.00
Hispanic 10th-grade enrollment										
≤ 3 percent	289	165	39.26	36.64	0.96	0.03	38.30	38.30	0.00	0.00
> 3 percent	463	304	60.74	63.36	-0.96	-0.02	61.70	61.70	0.00	0.00
Other[6] 10th-grade enrollment										
≤ 80 percent	365	235	48.37	48.38	-0.00	-0.00	48.37	48.37	0.00	0.00
> 80 percent	387	234	51.63	51.62	0.00	0.00	51.63	51.63	0.00	0.00

* Statistically significant at the 0.05/(c-1) level, where c is the number of categories within the primary variable.
[1] Design weight multiplied by the measure of size is used before nonresponse adjustment. This is the distribution to each response category.
[2] Weight after nonresponse adjustment multiplied by the measure of size is used.
[3] IEP = Individualized education program.
[4] LEP = Limited English proficient.
[5] Collapsed category comprising two Census divisions.
[6] Other includes all races/ethnicities other than Asian, Black, and Hispanic.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 21. Nonresponse bias before and after nonresponse adjustment for selected continuous variables for schools: 2002

	Before nonresponse adjustment						After nonresponse adjustment			
Description	Unweighted respondents	Unweighted non-respondents	Respondent mean weighted[1]	Non-respondent mean weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Relative bias
Minutes per class period	738	424	61.89	59.37	0.88*	0.01	61.01	60.71	0.30	0.01
Class periods per day	736	427	6.07	6.17	-0.03	-0.01	6.11	6.11	-0.01	-0.00
Total enrollment	744	461	1,229.56	1,398.30	-61.75*	-0.05	1,291.31	1,293.01	-1.70	-0.00
10th-grade enrollment	752	425	314.89	417.94	-35.47*	-0.10	350.36	329.98	20.38	0.06
Free or reduced-price lunch	685	395	23.86	19.25	1.61*	0.07	22.25	22.10	0.14	0.01
Number of full-time teachers	715	418	74.54	81.33	-2.39*	-0.03	76.93	77.38	-0.46	-0.01
Number of grades within the school	742	434	4.84	5.42	-0.20*	-0.04	5.042	5.094	-0.05	-0.01
IEP[3] percentage	697	407	9.10	8.18	0.32	0.04	8.78	8.79	-0.01	-0.00
LEP[4] percentage	713	411	3.68	5.70	-0.70*	-0.16	4.38	4.16	0.23	0.05
Number of part-time teachers	685	392	4.28	5.26	-0.36*	-0.08	4.59	4.44	0.15	0.03
Full-time teachers certified	708	412	89.92	89.80	0.04	0.00	89.88	89.59	0.29	0.00
Number of days in school year	735	424	179.47	178.53	0.33	0.00	179.14	179.23	-0.09	0.00
Is the school coeducational?	740	436	1.09	1.10	-0.00	-0.00	1.10	1.10	-0.00	-0.00
Type of grades within the school	742	434	2.74	2.65	0.03	0.01	2.71	2.70	0.01	0.01
School type	752	469	1.33	1.44	-0.04	-0.03	1.37	1.37	0.00	0.00
Geocode	752	469	5.59	6.10	-0.19	-0.03	5.78	5.74	0.04	0.01
Asian 10th-grade enrollment	752	469	4.66	6.20	-0.57*	-0.11	5.22	5.35	-0.13	-0.02
Black 10th-grade enrollment	752	469	15.72	10.41	1.95*	0.14	13.76	13.87	-0.11	-0.01
Hispanic 10th-grade enrollment	752	469	11.24	14.14	-1.07*	-0.09	12.31	12.07	0.24	0.02
Other[5] 10th-grade enrollment	752	469	68.35	69.20	-0.31	-0.00	68.67	68.68	-0.02	0.00

* Statistically significant at the 0.05/(c-1) level, where c is the number of categories within the primary variable.
[1] Design weight multiplied by the measure of size is used before nonresponse adjustment. This is the mean of the continuous variable.
[2] Weight after nonresponse adjustment multiplied by the measure of size is used.
[3] IEP = Individualized education program.
[4] LEP = Limited English proficient.
[5] Other includes all races/ethnicities other than Asian, Black, and Hispanic.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 22. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for students: 2002

	Before nonresponse adjustment						After nonresponse adjustments				
Description	Unweighted respondents	Unweighted non-respondents	Respondent mean weighted[1]	Non-respondent mean weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Bias / standard error	Relative bias
Minutes per class period											
≤ 45	3,595	596	18.56	23.35	-0.60*	-0.03	19.16	19.23	-0.07	-0.21	-0.00
46–50	3,247	414	21.85	19.45	0.30	0.01	21.55	21.58	-0.03	-0.06	-0.00
51–80	4,032	636	28.93	32.41	-0.43	-0.01	29.36	29.29	0.07	0.16	0.00
81+	4,274	494	30.66	24.79	0.73*	0.02	29.93	29.90	0.03	0.08	0.00
Class periods per day											
1–4	4,370	491	31.42	24.44	0.87*	0.03	30.55	30.63	-0.08	-0.23	-0.00
5–6	3,680	645	26.72	32.86	-0.77*	-0.03	27.49	27.48	0.01	0.01	0.00
7	4,083	565	24.56	24.3	0.03	0.00	24.52	24.42	0.10	0.30	0.00
8–9	2,975	425	17.31	18.40	-0.14	-0.01	17.44	17.47	-0.03	-0.08	-0.00
Total enrollment											
≤ 600	3,619	300	18.27	9.42	1.12*	0.07	17.15	17.17	-0.02	-0.04	-0.00
601–1,200	4,641	610	27.53	23.89	0.46	0.02	27.07	27.14	-0.07	-0.22	-0.00
1,201–1,800	3,455	545	26.09	28.87	-0.35	-0.01	26.44	26.40	0.04	0.11	0.00
>1,800	3,530	748	28.11	37.83	-1.23*	-0.04	29.34	29.30	0.04	0.08	0.00
10th-grade enrollment											
0–99	3,041	288	13.13	6.85	0.80*	0.06	12.35	12.35	0.00	0.00	0.00
100–249	3,976	433	22.55	16.27	0.80*	0.04	21.75	21.75	0.00	0.00	0.00
250–499	4,941	777	36.13	38.33	-0.28	-0.01	36.41	36.41	0.00	0.00	0.00
500+	3,404	731	28.16	38.55	-1.32*	-0.04	29.48	29.48	0.00	0.00	0.00
Free or reduced-price lunch											
0	2,677	338	8.64	7.88	0.09	0.01	8.55	8.58	-0.03	-0.18	-0.00
1–10	3,220	531	24.83	32.85	-0.98*	-0.04	25.80	25.69	0.11	0.33	0.00
11–30	4,141	492	35.82	29.74	0.74*	0.02	35.08	35.22	-0.14	-0.33	-0.00
> 30	4,063	568	30.71	29.53	0.14	0.00	30.57	30.51	0.06	0.13	0.00
Number of full-time teachers											
1–40	3,817	333	18.49	10.94	0.93*	0.05	17.56	17.44	0.12	0.42	0.01
41–70	3,822	455	23.97	20.26	0.46	0.02	23.51	23.51	0.00	0.00	0.00
71–100	3,667	619	29.36	32.92	-0.44	-0.01	29.80	29.99	-0.19	-0.45	-0.01
101+	3,328	644	28.17	35.88	-0.95*	-0.03	29.12	29.05	0.07	0.17	0.00
Number of grades within the school											
4	11,532	1,721	79.12	84.42	-0.66*	-0.01	79.79	79.70	0.09	0.26	0.00
> or < 4	3669	428	20.88	15.58	0.66*	0.03	20.21	20.30	-0.09	-0.26	-0.00
IEP[3] percentage											
≤ 5	5,600	744	26.41	26.87	-0.06	-0.00	26.47	26.52	-0.05	-0.15	-0.00
6–10	3,672	531	32.98	35.84	-0.35	-0.01	33.33	33.37	-0.04	-0.09	-0.00
11–15	3,139	347	26.71	20.21	0.79*	0.03	25.92	25.83	0.09	0.29	0.00
>15	1,943	351	13.90	17.09	-0.39	-0.03	14.28	14.28	0.00	0.00	0.00

See notes at end of table.

Table 22. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for students: 2002–Continued

	Before nonresponse adjustment						After nonresponse adjustments				
Description	Unweighted respondents	Unweighted non-respondents	Respondent mean weighted[1]	Non-respondent mean weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Bias / standard error	Relative bias
LEP[4] percentage											
0	6,609	749	36.42	28.19	1.00*	0.03	35.41	35.25	0.16	0.44	0.01
1	2,822	405	22.61	24.25	-0.20	-0.01	22.81	22.76	0.05	0.15	0.00
2–5	2,421	388	18.39	19.75	-0.17	-0.01	18.55	18.70	-0.15	-0.50	-0.01
>6	2,766	475	22.58	27.81	-0.64	-0.03	23.22	23.29	-0.07	-0.13	-0.00
Number of part-time teachers											
0–1	4,109	544	31.95	30.00	0.24	0.01	31.71	31.65	0.06	0.16	0.00
2–3	4,015	494	28.68	25.51	0.39	0.01	28.29	28.31	-0.02	-0.03	-0.00
4–6	3,345	459	20.85	19.08	0.22	0.01	20.63	20.63	0.00	0.00	0.00
7+	2,551	451	18.51	25.41	-0.85*	-0.04	19.36	19.42	-0.06	-0.16	-0.00
Full-time teacher certified											
0–90	3,569	521	15.52	16.88	-0.17	-0.01	15.69	15.59	0.10	0.25	0.01
91–99	2,565	335	20.01	18.91	0.13	0.01	19.87	19.93	-0.06	-0.17	-0.00
100	8,388	1140	64.47	64.21	0.03	0.00	64.44	64.47	-0.03	-0.08	0.00
Number of days in school year											
Less than 180	3,948	486	24.97	21.11	0.48	0.02	24.50	24.40	0.10	0.27	0.00
180	8,339	1,191	56.91	56.93	-0.00	0.00	56.91	56.84	0.07	0.17	0.00
More than 180	2,777	431	18.12	21.96	-0.47	-0.03	18.59	18.75	-0.16	-0.51	-0.01
Is the school coeducational?											
Yes	14,369	2,036	97.89	98.27	-0.05	-0.00	97.94	97.94	0.00	-0.14	0.00
No, all-female school	365	40	1.02	0.90	0.01	0.01	1.00	1.00	0.00	0.11	0.00
No, all-male school	420	50	1.09	0.83	0.03	0.03	1.06	1.06	0.00	0.10	0.00
Type of grades within the school											
K–12, PreK–10th, 1st–12th, PreK/1st – 9th/12th and PreK–12	998	118	5.21	2.88	0.29*	0.06	4.92	4.91	0.01	0.03	0.00
Middle grades but no elementary	1647	175	7.95	5.15	0.35*	0.05	7.60	7.59	0.01	0.12	0.00
Only high school	12,558	1,856	86.84	91.97	-0.64*	-0.01	87.48	87.51	-0.03	-0.07	0.00
School type											
Public	12,039	1,843	92.12	94.04	-0.24*	-0.00	92.36	92.36	0.00	0.00	0.00
Catholic	1,920	193	4.39	3.25	0.15*	0.03	4.25	4.25	0.00	0.00	0.00
Other private	1,403	193	3.49	2.71	0.10	0.03	3.39	3.39	0.00	0.00	0.00
Metropolitan status											
Urban	5,115	873	29.37	35.52	-0.78*	-0.03	30.15	30.15	0.00	0.00	0.00
Suburban	7,399	1,064	50.34	49.71	0.08	0.00	50.26	50.26	0.00	0.00	0.00
Rural	2,848	292	20.29	14.77	0.70*	0.04	19.59	19.59	0.00	0.00	0.00
Geocode											
Census division (public schools)											
Public—New England/Middle Atlantic[5]	2,021	489	15.55	24.45	-1.13*	-0.07	16.68	16.67	0.01	0.03	0.00
Public—East North Central	1,920	281	14.39	14.86	-0.06	-0.00	14.45	14.41	0.04	0.17	0.00

See notes at end of table.

Table 22. Nonresponse bias before and after nonresponse adjustment for selected categorical variables for students: 2002–Continued

	Before nonresponse adjustment						After nonresponse adjustments				
Description	Unweighted respondents	Unweighted non-respondents	Respondent mean weighted[1]	Non-respondent mean weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Bias / standard error	Relative bias
Public—West North Central	994	105	7.92	6.83	0.14	0.02	7.78	7.91	-0.13	-0.87	-0.02
Public—South Atlantic	2,236	316	16.56	15.53	0.13	0.01	16.43	16.29	0.14	0.54	0.01
Public—East South Central	888	78	6.30	3.92	0.30*	0.05	6.00	5.93	0.07	0.38	0.01
Public—West South Central	1,428	143	9.94	6.49	0.44*	0.05	9.50	9.53	-0.03	-0.14	-0.00
Public—Mountain	660	101	7.12	6.21	0.12	0.02	7.01	7.10	-0.09	-0.30	-0.01
Public—Pacific	1,892	330	14.34	15.76	-0.18	-0.01	14.52	14.51	0.01	0.02	0.00
Census Region (private schools)											
Private—Northeast	742	86	1.91	1.47	0.06	0.03	1.85	1.86	-0.01	-0.12	-0.01
Private—Midwest	983	112	2.00	1.73	0.03	0.02	1.97	1.88	0.09	1.75	0.05
Private—South	1,070	91	2.45	1.30	0.15*	0.06	2.30	2.47	-0.17	-2.28	-0.07
Private—West	528	97	1.52	1.46	0.01	0.01	1.51	1.43	0.08	0.39	0.05
Asian 10th-grade enrollment percent											
≤ 2 percent	5,963	818	38.50	37.08	0.18	0.00	38.32	38.32	0.00	0.00	0.00
> 2 percent	9,399	1,411	61.50	62.92	-0.18	-0.00	61.68	61.68	0.00	0.00	0.00
Black 10th-grade enrollment percent											
≤ 4 percent	5214	771	34.47	34.64	-0.02	-0.00	34.49	34.49	0.00	0.00	0.00
> 4 percent	10,148	1,458	65.53	65.36	0.02	0.00	65.51	65.51	0.00	0.00	0.00
Hispanic 10th-grade enrollment percent											
≤ 3 percent	5,974	788	37.99	35.96	0.26	0.01	37.74	37.74	0.00	0.00	0.00
> 3 percent	9,388	1,441	62.01	64.04	-0.26	-0.00	62.26	62.26	0.00	0.00	0.00
Other[6] 10th-grade enrollment percent											
≤ 80 percent	7,582	1,212	50.74	54.86	-0.52	-0.01	51.26	51.26	0.00	0.00	0.00
> 80 percent	7,780	1,017	49.26	45.14	0.52	0.01	48.74	48.74	0.00	0.00	0.00
Student sex											
Male	6,973	1,078	50.07	54.19	-0.52*	-0.01	50.58	50.60	-0.02	-0.05	0.00
Female	7,013	905	49.93	45.81	0.52*	0.01	49.42	49.40	0.02	0.05	0.00
Student race/ethnicity											
Asian	1,579	312	3.66	5.18	-0.19*	-0.05	3.85	3.94	-0.09	-1.19	-0.02
Black	2,019	304	15.61	16.63	-0.13	-0.01	15.74	15.82	-0.08	-0.30	-0.01
Hispanic	1,724	277	13.27	15.75	-0.30	-0.02	13.57	13.56	0.01	0.03	0.00
Other	8,803	1076	67.46	62.43	0.62*	0.01	66.84	66.68	0.16	0.42	0.00

* Statistically significant at the 0.05/(c-1) level, where c is the number of categories within the primary variable.
[1] Design weight is used before nonresponse adjustment. This is the distribution to each response category.
[2] Weight after nonresponse adjustment.
[3] IEP = Individualized education program.
[4] LEP =Limited English proficient.
[5] Collapsed category comprising two Census divisions.
[6] Other includes all races/ethnicities other than Asian, Black, and Hispanic.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 23. Nonresponse bias before and after nonresponse adjustment for selected continuous variables for students: 2002

	Before nonresponse adjustment						After nonresponse adjustments				
Description	Unweighted respondents	Unweighted non-respondents	Respondent mean, weighted[1]	Non-respondent mean, weighted[1]	Estimated bias	Relative bias	Overall mean, before adjustments[1]	Overall mean, after adjustments[2]	Estimated bias	Bias / standard error	Relative bias
Minutes per class period	15,148	2,140	63.01	60.70	0.29	0.00	62.72	62.67	0.05	0.33	0.00
Class periods per day	15,108	2,126	5.96	6.14	-0.02*	-0.00	5.98	5.98	0.00	0.13	0.00
Total enrollment	15,245	2,203	1,375.31	1,679.32	-38.47*	-0.03	1,413.78	1,408.91	4.87	0.58	0.00
10th-grade enrollment	15,362	2,229	368.44	455.93	-11.13*	-0.03	379.57	377.79	1.78	0.72	0.01
Free or reduced-price lunch	14,101	1,929	25.88	25.71	0.02	0.00	25.86	25.72	0.15	0.57	0.01
Number of full-time teachers	14,634	2,051	83.60	95.30	-1.44*	-0.02	85.04	85.18	-0.15	-0.40	-0.00
Number of grades within the school	15,201	2,149	4.57	4.30	0.03*	0.01	4.43	4.53	-0.09	0.18	-0.02
IEP[3] percentage	14,354	1,973	10.41	10.80	-0.05	-0.00	10.46	10.45	0.01	0.16	0.00
LEP[4] percentage	14,618	2,017	4.48	5.07	-0.07	-0.02	4.55	4.64	-0.09	-0.93	-0.02
Number of part-time teachers	14,020	1,948	4.07	5.03	-0.12*	-0.03	4.19	4.18	0.01	0.17	0.00
Full-time teacher certified	14,522	1,996	94.16	94.75	-0.07	-0.00	94.23	94.29	-0.06	-0.43	-0.00
Number of days in school year	15,064	2,108	179.26	180.27	-0.13	-0.00	179.38	179.36	0.03	0.44	0.00
Is the school coeducational?	15,154	2,126	1.03	1.03	0.00	0.00	1.03	1.03	0.00		0.00
Type of grades within the school	15,203	2,149	2.82	2.89	-0.01	-0.00	2.83	2.83	0.00		0.00
School type	15,362	2,229	1.11	1.09	0.00	0.00	1.11	1.11	0.00		0.00
Metropolitan status	15,362	2,229	1.91	1.79	0.01	0.01	1.89	1.89	0.00		0.00
Geocode	15,362	2,229	4.72	4.27	0.06	0.01	4.67	4.67	-0.00		0.00
Asian 10th-grade enrollment	15,362	2,229	4.23	5.66	-0.18*	-0.04	4.41	4.47	-0.05	-0.63	-0.01
Black 10th-grade enrollment	15,362	2,229	15.08	16.53	-0.18	-0.01	15.27	15.12	0.15	0.84	0.01
Hispanic 10th-grade enrollment	15,362	2,229	12.52	15.66	-0.40*	-0.03	12.92	12.81	0.11	0.45	0.01
Other race[5] 10th-grade enrollment	15,362	2,229	68.15	62.15	0.76*	0.01	67.39	67.60	-0.21	-0.70	-0.00
Student sex	13,986	1,983	1.50	1.46	0.01	0.00	1.49	1.49	0.00	0.05	0.00
Student race/ethnicity	14,125	1,969	3.45	3.36	0.01	0.00	3.43	3.43	0.00	0.67	0.00

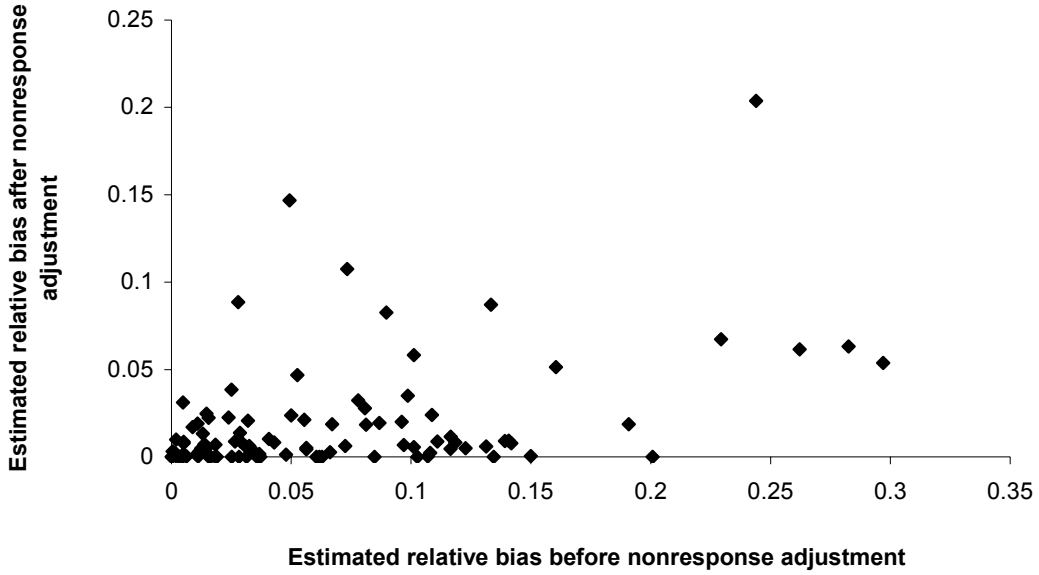
* Statistically significant at the 0.05/(c-1) level, where c is the number of categories within the primary variable.
[1] Design weight is used before nonresponse adjustment. This is the mean of the continuous variable.
[2] Weight after nonresponse adjustment.
[3] IEP = Individualized education program.
[4] LEP =Limited English proficient.
[5] Other race/ethnicity than Asian, Black, and Hispanic.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Figures 2 and 3 compare the estimated relative bias before nonresponse adjustment with the estimated relative bias after nonresponse adjustment for schools and students, respectively. Relative bias is the bias of the estimates divided by the estimate. It provides an indication of the order of magnitude of the bias with respect to the estimate. Both figures indicate that when the relative bias was large before nonresponse adjustment, it was almost always reduced dramatically after nonresponse adjustment. When the relative bias was small before nonresponse adjustment, it stayed small after nonresponse adjustment with occasional small increases. These figures clearly show that the nonresponse adjustment significantly reduced bias for schools and students.

Nonresponse bias can have an effect on significance testing. Table 21 includes an estimate of the bias ratio (student bias divided by the standard error). If this ratio is larger than 2 percent, then the probability of a Type I error is greater than 0.05. Figure 4 shows the student bias ratio by the Type I error rate. This figure shows that for most of the student variables included in the nonresponse bias analysis, the Type I error rate is 0.05, and two outliers were not graphed. This figure does not take the school bias ratio into account. The school bias ratio varies by school variable. If it is assumed that the school bias ratio is zero (the minimum value using the school-level nonresponse bias analysis variables), then there is no effect on the student bias ratio. However, if the school bias ratio is large (the maximum value using the school-level nonresponse bias analysis variables), then the Type I error rates are greater than 0.32. The data user should exercise caution when conducting statistical tests.

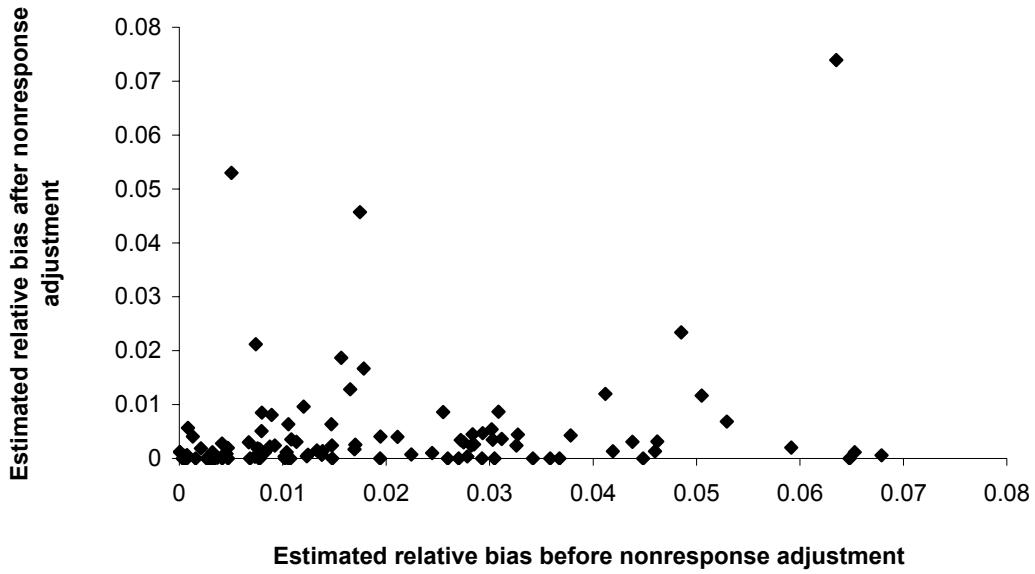
No additional nonresponse bias analysis was necessary to account for nonresponse from school administrators, libraries, or facility checklists because each of these had a response rate greater than 95 percent.

Figure 2. Before versus after nonresponse adjustment—school-level relative bias: 2002



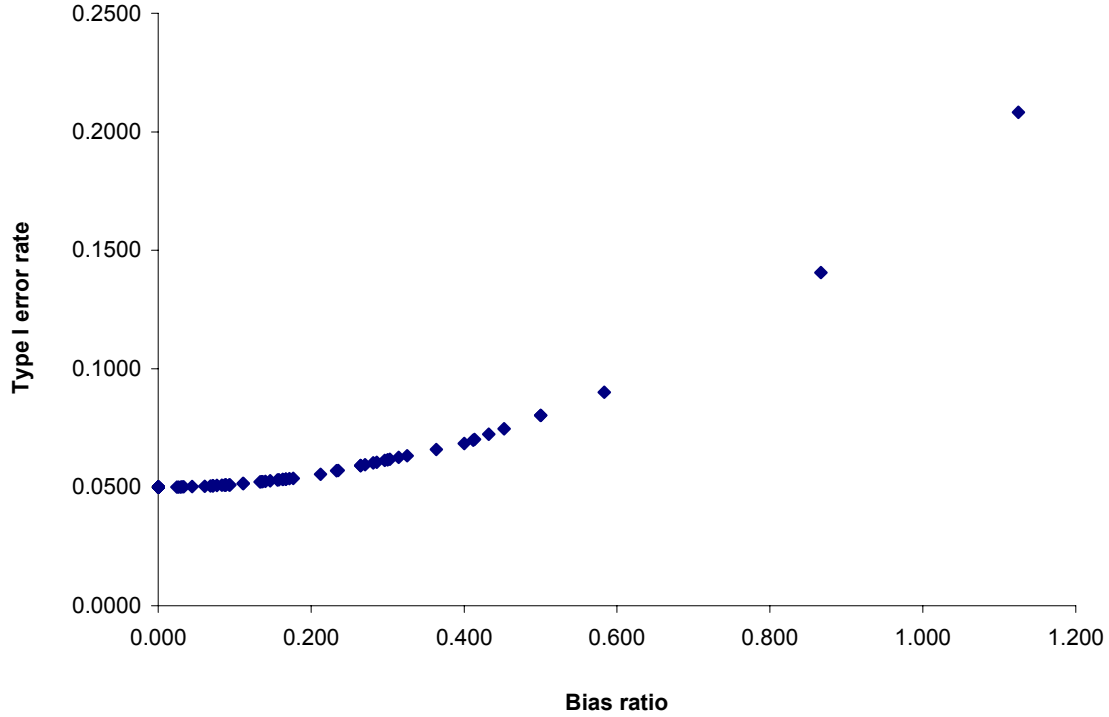
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Figure 3. Before versus after nonresponse adjustment—Student-level relative bias: 2002



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Figure 4. Minimum bias ratio by Type I error rate: 2002



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

In conclusion, examination of variables known for most respondents and nonrespondents before nonresponse adjustment detected some degree of bias. The school and student nonresponse bias analyses in conjunction with the weighting adjustments described in section 3.4 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 data user should exercise caution in using the data because bias was not estimated and corrected for all variables. The relative bias decreased considerably after weight adjustments—especially when it was large before nonresponse adjustment, and the relative bias remained small after weight adjustments when it was small before nonresponse adjustment usually.

As shown in figures 2 and 3, nonresponse bias was reduced using weighting techniques, and the remaining relative bias ranged from 0 to 0.2 percent for schools and from 0 to 0.07 percent for students.

3.2.7 Questionnaire Item Nonresponse Bias Analysis

This section (along with appendix I) documents the bias associated with item nonresponse, for the high nonresponse items on the ELS:2002 questionnaires. The NCES Statistical Standards[28] note that "nonresponse bias occurs when the observed value deviates from the population parameter due to differences between respondents and nonrespondents."

Data users are urged to take nonresponse bias into account, particularly when employing the high nonresponse variables described below. "High nonresponse" is defined as instances in which item response falls below the NCES standard of a minimum of 85 percent. Response rates are calculated in accordance with NCES Standard 1-3-5. Specifically, NCES Standard 1-3-5 stipulates that item response rates (RRI) are to be calculated as the ratio of the number of respondents for whom an in-scope response was obtained (I^x for item x) to the number of respondents who are asked to answer that item. The number asked to answer an item is the number of unit level respondents (I) minus the number of respondents with a valid skip for item x (V^x):

$$RRI^X = \frac{I^x}{I - V^x}$$

The ELS:2002 ECB data are housed in two megafiles, one at the student level (containing data from the student, parent and teacher questionnaires), and one at the school level (containing data from the school administrator and library media center questionnaires, and from the facilities checklist). For student-level estimates the final (i.e, nonresponse-adjusted) student weight (BYSTUWT) is used in the item response rate calculation. For school-level estimates, the final school weight (BYSCHWT) is employed in the calculation.

[28] See U.S. Department of Education, National Center for Education Statistics (2002). The statistical standards can also be accessed online at *http://nces.ed.gov/statprog/2002/stdtoc.asp*.

3.2.7.1 High Nonresponse Questionnaire Variables: Student-Level Items

No parent or teacher questionnaire items fell below 85 percent response. However, there were 78 such items on 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. Student-survey item nonresponse is primarily a function of questionnaire position, with the highest nonresponse seen in the final item, which was answered by only 64.6 percent of respondents. The 78 student variables evidencing high (>15 percent) nonresponse[29] are listed in table 24.

Table 24. Student-level high nonresponse questionnaire variables, by weighted response rate: 2002

Variable name	Description	Weighted item response rate
BYWORKSY	Student held job for pay during 2001–02 school year	84.3
BYS65B	How far in school father wants 10th grader to go	82.4
BYS73	Date last worked for pay	84.5
BYS74	Date started current/most recent job	76.9
BYS75	How many hours usually works a week	81.7
BYS76	How many hours works on the weekend	81.2
BYS77	Type of work does on current/most recent job	80.6
BYS79	How got current/most recent job	83.1
BYS80	How closely related job is to desired job after education	83.8
BYS85C	Special privileges given for good grades	84.6
BYS85D	Parents limit privileges due to poor grades	84.7
BYS85E	Required to work around the house	82.2
BYS85F	Parents limit TV watching or video games	84.6
BYS85G	Parents limit time with friends	83.8
BYS86A	How often discussed school courses with parents	82.7
BYS86B	How often discussed school activities with parents	82.5
BYS86C	How often discuss things studied in class with parents	82.1
BYS86D	How often discussed grades with parents	82.2
BYS86E	How often discussed transferring with parents	81.6
BYS86F	How often discussed prep for ACT/SAT with parents	81.9
BYS86G	How often discussed going to college with parents	81.6
BYS86H	How often discussed current events with parents	81.7
BYS86I	How often discussed troubling things with parents	81.4
BYS87A	Gets totally absorbed in mathematics	77.4
BYS87B	Thinks reading is fun	77.8
BYS87C	Thinks math is fun	77.0
BYS87D	Reads in spare time	76.6
BYS87E	Gets totally absorbed in reading	76.5
BYS87F	Mathematics is important	77.5
BYS88A	Most people can learn to be good at math	76.6
BYS88B	Have to be born with ability to be good at math	77.0
BYS89A	Can do excellent job on math tests	75.3
BYS89B	Can understand difficult math texts	75.6
BYS89C	Can understand difficult English texts	74.5

See note at end of table.

[29] For further details about these variables, see codebooks of response frequencies in appendix G and questionnaire facsimiles in appendix B. These appendices can be found in the electronic version of this user's manual on the NCES web site, in the form of a PDF file (*http://nces.ed.gov/surveys/els2002/*).

Table 24. Student-level high nonresponse questionnaire variables, by weighted response rate: 2002–Continued

Variable name	Description	Weighted item response rate
BYS89D	Studies to get a good grade	74.7
BYS89E	Can learn something really hard	73.6
BYS89F	Can understand difficult English class	74.5
BYS89G	Remembers most important things when studies	73.4
BYS89H	Studies to increase job opportunities	73.4
BYS89I	Can do excellent job on English assignments	72.6
BYS89J	Works as hard as possible when studies	73.5
BYS89K	Can do excellent job on English tests	72.4
BYS89L	Can understand difficult math class	73.0
BYS89M	Can master skills in English class	72.0
BYS89N	Can get no bad grades if decides to	72.7
BYS89O	Keeps studying even if material is difficult	71.7
BYS89P	Studies to ensure financial security	72.3
BYS89Q	Can get no problems wrong if decides to	71.3
BYS89R	Can do excellent job on math assignments	71.7
BYS89S	Does best to learn what studies	70.7
BYS89T	Can learn something well if wants to	71.2
BYS89U	Can master math class skills	70.5
BYS89V	Puts forth best effort when studying	71.2
BYS90A	Important to friends to attend classes regularly	70.4
BYS90B	Important to friends to study	71.0
BYS90C	Important to friends to play sports	69.8
BYS90D	Important to friends to get good grades	70.3
BYS90E	Important to friends to be popular with students	69.4
BYS90F	Important to friends to finish high school	70.1
BYS90G	Important to friends to have steady boy/girlfriend	69.2
BYS90H	Important to friends to continue education past high school	69.8
BYS90J	Important to friends to do community work	69.7
BYS90K	Important to friends to have job	68.8
BYS90L	Important to friends to get together with friends	69.6
BYS90M	Important to friends to go to parties	68.5
BYS90Q	Important to friends to make money	68.3
BYS91	Number of close friends who dropped out	66.8
BYS92A	Girls should have same opportunities in sports	67.2
BYS92B	Some sports should be just for boys	67.6
BYS92C	Girls should have own sports teams	66.3
BYS92D	Girls should be on same sports teams as boys	67.6
BYS94	Has close friends who were friends in 8th grade	65.0
BYS96	Observed students betting on sports	64.7
BYS97A	Bets were placed with friends	64.6
BYS97B	Bets were placed with family members	64.6
BYS97C	Bets were placed with bookie	64.6
BYS97D	Bets were placed with a website	64.6
BYS97E	Bets were placed through other means	64.6

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.2.7.2 High Nonresponse Questionnaire Variables: School-Level Items

At the school level, 41 administrator items fell below 85 percent (ranging from 84.7 percent to a low of 74.6 percent). No library media center questionnaire items fell below the threshold, nor did any facility checklist items fall below 85 percent. While the school-level items will 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, and second, the fact that some administrators completed an abbreviated version of the school administrator questionnaire (the high nonresponse items did not appear on the abbreviated instrument).

Forty-one school-level questionnaire variables evidencing high (>15 percent) nonresponse are listed in table 25:

Table 25. School-level high nonresponse questionnaire variables, by weighted response rate: 2002

Variable name	Description	Weighted item response rate
BYA14A	Percent 10th graders in general high school program	84.7
BYA14C	Percent 10th graders in other specialized programs	82.1
BYA14F	Percent 10th graders in alternative program	83.1
BYA14G	Percent 10th graders receive bilingual education	82.8
BYA14H	Percent 10th graders receive ESL	84.7
BYA14I	Percent 10th graders receive remedial reading	83.8
BYA14J	Percent 10th graders receive remedial math	83.8
BYA14K	Percent 10th graders in after school/summer outreach	81.2
BYA23C	Number of full-time art teachers	81.9
BYA23F	Number of full-time foreign language teachers	81.8
BYA23I	Number of full-time vocational education teachers	81.8
BYA23J	Number of full-time physical education teachers	83.7
BYA23L	Number full-time special education teachers	83.6
BYA24B	Percent part-time teachers are certified	81.2
BYA25A	Percent full-time teachers teach out of field	84.3
BYA25B	Percent part-time teachers teach out of field	75.7
BYA26A	Lowest salary paid to full-time teachers	81.4
BYA26B	Highest salary paid to full-time teachers	81.2
BYA30	Main source of content standards	80.2
BYA33CA	Minimum competency test given in grade 9	83.0
BYA33CB	Math is on grade 9 competency test	81.9
BYA33CC	Science is on grade 9 competency test	81.9
BYA33CD	English is on grade 9 competency test	81.9
BYA33CE	History/social studies is on grade 9 competency test	81.9
BYA33EA	Minimum competency test given in grade 11	83.4
BYA33EB	Math is on grade 11 competency test	83.0
BYA33EC	Science is on grade 11 competency test	83.0
BYA33ED	English is on grade 11 competency test	83.0

See note at end of table.

Table 25. School-level high nonresponse questionnaire variables, by weighted response rate: 2002–Continued

Variable name	Description	Weighted item response rate
BYA33EE	History/social studies is on grade 11 competency test	83.0
BYA33FA	Minimum competency test given in grade 12	81.8
BYA33FB	Math is on grade 12 competency test	81.5
BYA33FC	Science is on grade 12 competency test	81.5
BYA33FD	English is on grade 12 competency test	81.5
BYA33FE	History/social studies is on grade 12 competency test	81.5
BYA47A	School's relationship with school board	83.8
BYA47B	School's relationship with central office	79.8
BYA47C	School's relationship with teachers' association	74.6
BYA48E	Principal evaluated on relationship with community	84.1
BYA48F	Principal evaluated on new programs/reform	83.8
BYA50F	Learning hindered by poor library	84.6
BYA50K	Learning hindered by poor voc/tech equipment/facilities	84.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.2.7.3 Estimating Bias

The potential magnitude of nonresponse bias can be estimated by taking the product of the nonresponse rate and the difference in values of a characteristic between respondents and nonrespondents.

The possibility of estimating the degree of bias depends on having some variables that reflect key characteristics of respondents and for which there is little or no missing data. According to the NCES statistical standards, statistically-imputed data cannot be used for this purpose. This requirement leaves a limited number of characteristics that can be employed to help estimate the magnitude of bias. One source of relevant markers of the sample that meets the high response criterion can be found in frame variables from which the school sample was selected. The following such variables have therefore been incorporated into the bias analysis: school type (public, Catholic, other private); region (North, South, Midwest, West); and metropolitan status or urbanicity (urban, suburban, rural). These three variables (or ten characteristics) have been used to generate both student- and school-level analyses. For all ten characteristics, coverage is 100 percent.

In addition, a few key student classification variables have extremely high response rates. Therefore, these variables have been employed in the student-level bias analysis. These include sex (99.95 percent complete) and race/ethnicity (99.98 percent complete). These variables have also been included in the analysis. Other variables that have been included are: mother's education (96 percent coverage), language minority status (98 percent complete), reading quartile (high, middle two, low) (94 percent complete), and math quartile (high, middle two, low) (95 percent complete).

Despite the limitations imposed by some missingness, these variables are hypothesized to be especially helpful in explaining student questionnaire nonresponse and its biases. Since, on the student questionnaire, nonresponse is primarily a function of question position, one may hypothesize that poor readers in particular (or poor students more generally) are most likely to be

missing on the final items. An additional reason for including both the math and reading quartile is that a number of the high nonresponse student variables have to do with psychological orientations toward mathematics or English, such that, for these scales, any bias by reading or mathematics achievement level would be particularly interesting to quantify.

3.2.7.4 Summarization of Bias: Magnitude of Bias; Characteristics Associated with Bias

Appendix I[30] contains tables listing all high nonresponse variables. For the student questionnaire, there were 78 such variables, and 40 relevant characteristics (sex [male or female], race/ethnicity [seven categories], mother's education [eight levels], school sector [public, Catholic, or other private], metropolitan status of school locale [three levels of urbanicity: urban, suburban, rural], Census region [North, South, Midwest, West], reading quartile [highest quartile, middle quartiles, lowest quartile], math quartile [highest quartile, middle quartiles, lowest quartile] and home language [English or non-English]).[31] Thus there are (in total) 3,120 observations. For all observations, appendix I provides the signed bias estimate, and *t* values for tests of whether the estimate differs significantly from zero at .05 (*t* must be 1.96 or higher to meet this probability criterion).

For the school administrator questionnaire, there were 41 high nonresponse variables (< 85 percent) and 10 characteristics (school sector [public, Catholic, or other private], metropolitan status of school locale [locale has three levels of urbanicity: urban, suburban, rural], and Census region [North, South, Midwest, West]). This yields 410 observations.

Tables 26-28 below summarize student-level bias for the 78 student questionnaire high nonresponse items.

Table 26. ELS:2002 student file, 78 high nonresponse variable summary, by 40 characteristics: mean, median, and standard deviation of bias estimates: 2002

Overall mean	1.20
Median	0.75
Standard deviation	1.19

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 27. Frequency distribution of unsigned bias estimates, 78 high nonresponse student variables, by 40 characteristics: 2002

Percentage range of bias	Frequency	Percent
0 <= Bias percent < 1.0	1725	55.29
1 <= Bias percent < 2.0	680	21.79
2 <= Bias percent < 3.0	364	11.67
3 <= Bias percent < 4.0	284	9.10
4 <= Bias percent < 5.0	53	1.70
Bias percent >= 5.0	14	0.45

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

[30] Appendix I can be found in the electronic version of this user's manual on the NCES web site, in the form of a PDF file (*http://nces.ed.gov/surveys/els2002/*).

[31] In addition to 35 characteristics, for four imputed variables, a holding category of "imputed" is also supplied (sex, race, mother's educational attainment, and home language). In a fifth instance, two Hispanic categories (race specified/not specified) are combined for one overall Hispanic ethnicity category.

Table 28. Mean, median, and standard deviation for bias estimates for each of 40 characteristics, across 78 high nonresponse student file variables: 2002

Characteristic	Mean	Median	Standard deviation
Sex			
Male	2.18	2.30	0.29
Female	2.17	2.29	0.29
Imputed value	0.01	0.01	0.01
Race/ethnicity			
American Indian	0.15	0.15	0.05
Asian	0.06	0.05	0.03
Black	2.56	2.72	0.57
Multiracial	0.07	0.07	0.05
White	4.42	4.60	0.74
Hispanic ethnicity	1.55	1.55	0.19
Imputed value	0.19	0.20	0.03
Mother's education attainment			
No high school diploma	1.38	1.42	0.24
High school graduate	0.11	0.09	0.10
2-year school, no degree	0.18	0.20	0.09
2-year school, degree	0.14	0.12	0.09
4-year school, no degree	0.15	0.13	0.09
4-year degree	1.17	1.30	0.26
Master's degree	0.46	0.43	0.13
Ph.D. degree	0.14	0.14	0.05
Imputed value	0.79	0.75	0.13
School sector			
Public	0.37	0.37	0.06
Catholic	0.45	0.45	0.04
Other private	0.09	0.08	0.04
School locale			
Urban	1.61	1.70	0.27
Suburban	1.48	1.52	0.37
Rural	0.15	0.12	0.11
School region			
Northeast	0.71	0.80	0.24
Midwest	0.55	0.56	0.09
South	2.50	2.76	0.83
West	1.24	1.33	0.59
Reading achievement			
Low quartile	3.37	3.38	0.32
Medium 2 quartiles	0.45	0.33	0.41
High quartile	3.09	3.30	0.52
Math achievement			
Low quartile	3.15	3.21	0.29
Medium quartile	0.45	0.37	0.35
High quartile	3.08	3.38	0.64
Home language			
No (non-English)	1.06	1.09	0.14
Yes (English)	2.82	2.81	0.15
Imputed value	1.76	1.76	0.07

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 28 illustrates the 40 characteristics used to analyze bias across the 78 student-questionnaire high-nonresponse variables. As the table shows, the bias estimate was less than 1 percent over half the time, and less than two percent 77 percent of the time. A bias of 5 percent or higher was detected in less than one half of one percent of the observations.

The characteristic associated with the highest mean bias across all the high nonresponse student questionnaire items was being a White sophomore (mean bias was 4.4 percent). The next factor in order of significance was falling in the lowest reading quartile (3.4 percent mean bias), followed by falling in the lowest math quartile (3.2 percent mean bias). The fourth and fifth factors were falling in the highest reading or highest math quartile (both had a mean bias of about 3.1 percent).

While Table 28 reports unsigned (non-directional) bias estimates, appendix I reports the direction of bias in relation to the population parameter (with a minus sign ["-"] or an implicit plus sign) for each observation. Since, for the five factors noted above, the sign is consistent across all observations, more can be said to interpret the relationship between bias and the five characteristics. Specifically, Whites were disproportionately likely to answer the high nonresponse questionnaire items, as were students in the highest math or highest reading quartile. On the other hand, sophomores in the lowest math or lowest reading quartile were the groups most likely to be item nonrespondents.

This pattern suggests that poor readers, in particular, and students with low tested achievement in reading or mathematics, generally, were the least likely to respond to high nonresponse items, presumably in part because they were unable to complete the student questionnaire within the set time limits.

A further point of interest is how often the bias estimate was statistically significant (different from 0 at .05). As can be confirmed from the *t*-values reported in appendix I, 946 observations proved to be statistically significant, or about 30 percent of the total (3,120 observations).

Tables 29-31 below summarize school level bias for the 41 school administrator questionnaire high nonresponse items.

Table 29. ELS:2002 school file, 41 high nonresponse variable summary, by 10 characteristics: mean, median, and standard deviation of bias estimates: 2002

Overall mean	1.12
Median	0.84
Standard deviation	1.11

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 30. Frequency distribution of unsigned bias estimates, 41 high nonresponse school variables, by 10 characteristics: 2002

Range of bias	Frequency	Percent
0 <= Bias percent < 1.0	234	57.07
1 <= Bias percent < 2.0	109	26.59
2 <= Bias percent < 3.0	52	12.68
3 <= Bias percent < 4.0	8	1.95
4 <= Bias percent < 5.0	3	0.73
Bias percent >= 5.0	4	0.98

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 31. Mean, median, and standard deviation for bias estimates for each of 10 characteristics across 41 high nonresponse school file variables: 2002

Overall bias summary by characteristic	Mean	Median	Standard deviation
School sector			
Public	2.03	1.52	2.00
Catholic	0.37	0.34	0.35
Other private	1.80	1.64	1.62
School locale			
Urban	1.65	1.65	0.73
Suburban	1.33	1.47	0.70
Rural	0.77	0.50	0.83
School region			
Northeast	1.10	0.94	0.55
Midwest	0.59	0.62	0.42
South	1.13	1.02	0.63
West	0.45	0.28	0.50

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

As may be seen, the bias estimate across the 410 observations was less than 1 percent some 57 percent of the time. Bias is less than 2 percent some 84 percent of the time. Only about 1 percent of observations show a bias of 5 percent or higher; less than 2 percent show a bias of 4 percent or higher. Overall, the highest mean bias was for public schools (a bias of 2.0 percent), followed by other private (a bias of 1.8 percent) and urban schools (a bias of 1.6 percent).

Across the 410 observations, three of the bias estimates (less than 1 percent) were significantly different from zero at a .05 level of probability. The three significant observations were BYA47C ("How would you characterize your school's relationship with…[the] teacher's association or union") by each of the three school control types (public, Catholic, other private). While no other bias estimates proved to be statistically significant at the school level, readers are cautioned that this provides no assurance that the same result would necessarily hold for these variables at the student level of analysis. Indeed, the larger sample sizes at the student level would necessarily increase the tendency of bias estimates to differ significantly from zero. Additional caution is therefore advised in using these data when the student data are employed as the unit of analysis and the high-nonresponse school variables attached to the student level as a contextual extension. It should also be noted that nonresponse rates for these variables may differ at the student and school levels, owing to factors such as the variability in the number of students associated with each school administrator.

3.2.8 Other Data Quality Issues: Reports of Hours Spent in Viewing Television and Playing Video Games

Results obtained from analysis of data from the ELS:2002 base year generally conformed to expectations based on external sources and on theoretically established relationships between variables. However, a possible exception that is notable may be seen in the estimates for time spent watching television, videotapes or DVDs, and playing video games. While the general pattern of relationships conforms to past findings, the total number of hours registered was higher than expected. The paragraphs below provide more information about this data quality issue.

Students were asked to report the number of hours per day during the school year that they usually spent watching television, videotapes or DVDs (question 48) and playing video or computer games (question 49). Students were to write in a numerical value in hours per day within a constrained field, corresponding to the total number of hours watched (or played) per day on weekdays and, separately, on weekends.

Even after topcoding to eliminate implausibly extreme values, high-end estimates (proportion of the population engaged in television viewing over 5 or 6 hours per day) remained substantially higher than estimates from alternative sources such as NELS:88 or NAEP. There are a number of possible explanations for such a discrepancy. The two most important such explanations are (1) a lack of full comparability between sources and (2) the possibility that the ELS:2002 item was prone to misinterpretation by respondents who did not read it carefully.

Comparison with the NAEP television item (Campbell, Hombo, and Mazzeo 2000) is compromised by a number of factors. Over time there is fluctuation in estimates for the NAEP trend sample, which in any case is based on 13- and 17-year-olds (most ELS:2002 sophomores are 15 or 16 years of age). Moreover, the ELS:2002 item is broader, including additional viewing (specifically videotapes and DVDs) beyond television. The ELS:2002 item is open ended and elicits an answer that is continuous in form. In contrast, the NAEP item is categorical, with a tight cap on the highest response.

Comparison with NELS:88 (Rasinski, Ingels, Rock, and Pollack 1993) is also compromised by key differences, including a 12-year time gap and the fact that NELS:88 asked the item in categorical form. ELS:2002 is continuous. Estimates collected in an open-ended continuous format may differ from estimates collected in a constrained categorical format. The open-ended format may be more cognitively taxing, while the categorical format may influence response by implicitly defining the "comfortable" middle ranges as well as both extremes for respondents (Tourangeau, Rips, and Rasinski 2000). (For example, in NELS:88, respondents were asked to choose from response categories such as "less than 1 hour/day, 1–2 hours, 2–3 hours, 3–4 hours, and over 5 hours a day.")

Apart from the caveats that must be entered about the comparability of the ELS:2002 item, it is also important to consider that the ELS:2002 format may have been open to misinterpretation by some respondents. (This observation is speculative; it is not based on cognitive interviews with 10th graders, or re-interviews of ELS:2002 respondents.) In particular, although the question stems say, "how many hours a day," splitting the response boxes into

weekdays and weekends may have abetted some respondents in the error of reporting total weekday and total weekend hours. If some students forgot the definition in the question stem ("how many hours per day") and misinterpreted "weekdays" as the total number of hours on weekdays in a week, an inflated estimate for high-end use would be the likely consequence. A parallel error could be made for the "on weekends" portion of the question. Estimates from television viewing items in the past have been quite sensitive to small format differences (see Rasinski, Ingels, Rock, and Pollack 1993, appendix B, pp. 15–18). While reliable comparison sources are not available for the video game item, one may presume that because it was identical in format to the television viewing item, it would be open to a like degree of respondent error, and that that error would be in the same direction (i.e., somewhat inflated high-end estimates).

3.3 Imputation

The ELS:2002 data files contain school-level and student-level data collected from school administrator and teacher, parent, and student interviews, as well as from student assessments. These data were coded and edited to reflect skip-pattern relationships and different types of missing data. After the editing process was completed, the remaining missing values for 14 key analysis variables (see table 32) were imputed statistically. These variables were chosen because they are the row variables in the ELS:2002 *A Profile of the American High School Sophomore in 2002*. Most of the analysis variables were imputed using a weighted sequential hot deck procedure.[32] In addition, two further analysis variables,[33] ability estimates (theta) in mathematics and reading, were imputed using multiple imputation. The imputations were performed primarily to reduce the bias of survey estimates caused by missing data. Table 32 lists the variables in the order in which the missing data were imputed. The order of imputation addresses problems of multivariate association by using a series of univariate models fitted sequentially such that variables modeled earlier in the hierarchy had a chance to be included in the covariate set for subsequent models. Generally, the order of imputation for all variables was from the lowest percent missing to the highest. The percentage of missingness for each variable imputed is shown in table 32.

Before using the weighted sequential hot deck procedure, we imputed student sex logically. Logical imputation is a process that tries to determine whether the missing answer can be either deduced or guessed from answers to other questions. A distribution of student names by sex was used to impute student sex. Additionally, student race was logically imputed using student name and school-level information.

Sequential hot deck imputation is a common procedure used for item nonresponse. This method uses the respondent survey data (donors) to provide imputed values for records with missing values. The basic principle of sequential hot deck imputation involves defining imputation classes, which generally consist of a cross-classification of covariates, and then replacing missing values sequentially from a single pass through the survey data within the imputation classes. When sequential hot deck imputation is performed using the sampling weights (see section 3.4) of the item respondents and nonrespondents, the procedure is called

[32]See Cox (1980).
[33]Ability estimates (theta) are the precursors or bases for both the normative and criterion-referenced scores. By imputing theta, it was therefore possible to have 100 percent coverage for all test variables used in analysis.

weighted sequential hot deck imputation. This procedure takes into account the unequal probabilities of selection in the original sample by using the sampling weight to specify the expected number of times a particular respondent's answer will be used to replace a missing item. These expected selection frequencies are specified so that, over repeated applications of the algorithm, the expected value of the weighted distribution of the imputed values will equal in expectation within imputation class the weighted distribution of the reported answers.

Table 32. ELS:2002 imputation variables: 2002

Variable	Weighted percent missing
Student sex	0.05
Student race/ethnicity	0.02
Student language minority status	2.07
Student Hispanic subgroup	2.93
Student Asian subgroup	7.26
School program type	6.64
Student postsecondary educational aspirations	2.36
Parental aspirations for student postsecondary achievement	14.53
Family composition	12.55
Mother's educational attainment	3.88
Mother's occupation	5.58
Father's educational attainment	10.28
Father's occupation	15.03
Family income	22.40
Student ability estimates (theta) for reading	6.26
Student ability estimates (theta) for mathematics	5.33

NOTE: Additional reading and mathematics assessment variables generated on basis of imputed theta score.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Weighted sequential hot deck imputation has as an advantage that it controls the number of times that a respondent record can be used for imputation and gives each respondent record a chance to be selected for use as a hot deck donor. To implement the weighted hot deck procedure, imputation classes and sorting variables that were relevant for 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.

Multiple imputation is a technique that requires imputing missing values several times and creating *m* complete datasets. These datasets are created such that regular complete-case analyses can be performed. The parameters of interest, then, can be calculated by averaging the parameter estimators from each augmented data set. The SAS PROC MI procedure was used to impute two analysis variables (student ability estimates in reading and mathematics). The Markov Chain Monte Carlo (MCMC) model option in the SAS procedure, which assumes the data are from a multivariate normal distribution, was used to estimate the entire joint posterior probability distribution of the unknown quantities. Random draws from this distribution were taken to fill in the missing values. The variables included in the model were student race and sex, student language, student postsecondary aspirations, parental aspirations for student, family composition, mother's and father's occupation and education level, household income, and several school-level variables including school type, urbanicity, and census region.

Appendix F further documents the imputations performed. Table F-1 restates the imputation variables. Table F-2 presents the imputation classes and sorting variables used for all of the variables imputed by the weighted sequential hot deck approach, and table F-3 presents the variables used in the multiple imputation model. Table F-4 presents the before and after imputation distributions. To evaluate the effects of imputation, the distribution of variables was tested for significant differences. Statistical tests (*t* tests) were used to test each level of the variables for differences at the 0.05/(c-1) significance level, where c is the number of categories within the variable. Chi-squared tests were performed to test for significant differences in the distributions of each variable. Many of the test imputation variables showed significant differences at each level of the variable; however, the differences were usually small. Following data imputations, variables were reviewed and revised (if necessary) to adjust for inconsistencies with other known data.

3.4 Calculation of Weights and Results of Weighting

Three sets of weights were computed: a school weight, a weight for student questionnaire completion, and a contextual data weight for the "expanded" sample of questionnaire-eligible and questionnaire-ineligible students.

3.4.1 School Weights

School weights were computed in several steps. First, the school-level design weight (WT1) was calculated equal to the reciprocal of the school's probability of selection, which was:

$$\pi_r(i) = \begin{cases} \dfrac{n_r S_r(i)}{S_r(+)} & \text{for non-certainty selections} \\ 1 & \text{for certainty selections.} \end{cases}$$

where:

n_r = the sample size in stratum r,

$S_r(i)$ = the measure of size for the i-th school in stratum r, and

$S_r(+)$ = the total measure of size for all schools in stratum r.

Therefore, the school sampling weight was assigned as follows:

$WT1 = 1 / \pi_r(i)$.

For schools that were selected as part of the new school supplemental sample (see section 3.2.2), the probability of selection and WT1 were computed in the same manner. The values of n_r and $S_r(+)$ were the values for stratum r used to select the main sample.

Second, the school's design weight was adjusted to account for field test sampling. Field test schools were selected using stratified simple random sampling, and field test sample schools were then deleted from the full-scale school frame. To avoid compromising population coverage, each school on the full-scale sampling frame was assigned a first-stage sampling weight (WT2), which was the inverse of the probability of not being selected for the field test.

The schools in stratum r on the school sampling frame were partitioned as follows:

- Let i = 1, 2, …, $I_1(r)$ represent those schools not on the frame from which the field test sample was selected (new schools on the CCD and PSS).
- Let i=$I_1(r)$+1, $I_1(r)$+2, …, $I_2(r)$ represent those that were on the frame for the field test but were not selected.
- Let i=$I_2(r)$+1, $I_2(r)$+2, …, I(r) represent the schools in the simple random sample of n_{fr} schools selected for the field test.

The sampling weight component for the full-scale study was the reciprocal of the probability of not being selected for the field test, that is, for the i-th school in stratum r it was:

$$WT2_r(i) = \left\{ \begin{array}{cc} 1 & \text{for } i = 1,\ldots,I_1(r) \\ \dfrac{I(r) - I_1(r)}{I(r) - I_1(r) - n_{fr}} & \text{for } i = I_1(r)+1,\ldots,I_2(r) \end{array} \right\}$$

Third, the school weight was adjusted to account for the probability of the school being released. As described in section 3.2.2, a sample of 1,644 schools was selected and released in waves, as necessary, to achieve yield and response rate goals. However, not all schools were released, so the inverse of the school's probability of being released within a school stratum was taken to get a release weight (WT3).

WT3 = n_r / n_{re}, where n_{re} = number of schools released in stratum r.

It was assumed that all nonresponding schools were eligible because they were contacted repeatedly and there was no reason to believe that they were ineligible. Therefore, no adjustments were made for unknown school eligibility.

Next, generalized exponential models (GEMs)[34] which are a unified approach to nonresponse adjustment, poststratification, and extreme weight reduction were used. GEMs are a general version of weighting adjustments, and were based on a generalization of Deville and Särndal's logit model.[35] GEMs were a formalization of weighting procedures such as nonresponse adjustment, poststratification, and weight trimming. GEMs control at the margins as opposed to controlling at the cell level, as weighting class adjustments. This allows more variables to be considered.

A responding school was defined as a school that had a Survey Day. There were 752 such schools.[36] A list of all variables that were for the generality of both responding and nonresponding schools (these variables, some taken from the sampling frame and others collected from schools and districts, are listed in section 3.2.6) were compiled. 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. Then, these variables were main effects in the model. The weight specified for the

[34] See Folsom and Singh (2000).
[35] See Deville and Särndal (1992).
[36] One eligible school had no eligible students selected in the sample. This school was considered a responding school.

model was the product of WT1, WT2, and WT3 multiplied by the school's composite measure of size. The purpose of doing this was to account for the students at the first stage of sampling, because the students are the primary unit of analysis. The sample of schools was chosen with PPS and the measure of size was based on the number of students (see section 3.2.2). These variables were also used in Automatic Interaction Detection (AID) analyses (with response as the dependent variable) to determine important interactions for the model. The outcome of this first model was a nonresponse adjustment factor (WT4). The unequal weighting effects (UWEs) and maximum adjustment factors were monitored to ensure reasonable values. Table 33 presents the final predictor variables used in the nonresponse adjustment model and the average weight adjustment factors resulting from these variables. The nonresponse adjustment factors met the following constraints:

- minimum: 1.00
- median: 1.43
- maximum: 2.99

Quality control (QC) checks were performed on the weights as described in section 3.4.3. GEMs were designed so that the sum of the unadjusted weights for all eligible units equaled the sum of the adjusted weights for respondents. GEMs also constrained the nonresponse adjustment factors to be greater than or equal to one.

The innovation introduced in GEMs is the ability to incorporate specific lower and upper bounds. An important application of this feature is to identify at each adjustment step an initial set of cases with extreme weights and to use specific bounds to exercise control over the final adjusted weights. Thus, there is built-in control for extreme weights in GEM. Controlling extreme weights in this manner does not reduce the bias reduction potential of the adjustments. No extreme school weights needed trimming.

The primary purpose of the school weight is to be an intermediate weight for computing the student weight. However, some analysts are interested in doing school level analyses. While the school sample is a representative sample of schools in the target population, the school PPS selection was designed based on the number of students in various race/ethnicity categories. Therefore, the school weight, which takes the measure of size into account in the nonresponse adjustments, was not adequate for school level analyses. After comparing the school weights to CCD and PSS school counts, it was decided to compute a separate school weight that is included on the analysis file for school-level analyses. The intermediate school weight was used in computation of the student weight.

Table 33. Average weight adjustment factors used to adjust school weights for nonresponse: 2002

Model predictor variables[1]	Number of responding schools	Weighted response rate[2]	Average weight adjustment factor
Total	752	67.65	1.61
School type			
Public schools	580	68.93	1.59
Catholic schools	95	73.47	1.46
Other private schools	77	62.75	1.98
Metropolitan status			
Urban	250	67.05	1.64
Suburban	361	59.80	1.68
Rural	141	79.18	1.40
10th-grade enrollment			
0–99 10th-grade students	160	70.53	1.53
100–249 10th-grade students	187	65.43	1.52
250–499 10th-grade students	240	64.76	1.59
> 500 10th-grade students	165	56.07	1.83
Type of grades within school			
K–12, PreK–10th, 1st–12th, PreK/1st–9th/12th, and PreK–12 schools	57	67.03	2.21
Schools that contain middle grades but not elementary grades	79	68.66	1.41
Schools that only contain high school grades	616	67.73	1.58
Number of grades within the school			
4 grade levels within the school	53	69.55	1.44
> or < 4 grade levels within the school	699	67.58	1.62
Number of days in school year			
Less than 180 days of school in the school year	191	63.57	1.67
180 days of school in the school year	423	69.64	1.62
More than 180 days of school in the school year	138	66.89	1.50
Minutes per class period			
≤ 45 minutes per class period	178	66.78	1.63
46–50 minutes per class period	164	66.67	1.70
51–80 minutes per class period	198	68.62	1.71
≥ 81 minutes per class period	212	68.67	1.43
Class periods per day			
1–4 class periods per day	216	69.21	1.45
5–6 class periods per day	183	54.75	1.87
7 class periods per day	208	70.91	1.58
8–9 class periods per day	145	73.19	1.56
IEP[3] percentage			
≤ 5 percent IEP	303	64.86	1.56
6–10 percent IEP	190	66.06	1.80
11–15 percent IEP	158	72.39	1.59
> 15 percent IEP	101	73.04	1.42

See notes at end of table.

Table 33. Average weight adjustment factors used to adjust school weights for nonresponse: 2002–Continued

Model predictor variables[1]	Number of responding schools	Weighted response rate[2]	Average weight adjustment factor
LEP[4] percentage			
0 percent LEP	333	70.01	1.52
1 percent LEP	146	62.38	1.54
2–5 percent LEP	127	64.69	1.60
> 6 percent LEP	146	63.67	1.89
Free or reduced price lunch			
0 percent free or reduced-price lunch	141	53.40	1.85
1–10 percent free or reduced-price lunch	163	67.79	1.63
11–30 percent free or reduced-price lunch	222	70.19	1.58
> 31 percent free or reduced-price lunch	226	76.24	1.48
Number of full-time teachers			
1–40 full-time teachers	206	70.31	1.45
41–70 full-time teachers	189	61.67	1.58
71–100 full-time teachers	185	64.06	1.67
> 100 full-time teachers	172	68.36	1.78
Number of part-time teachers			
0–1 part-time teachers	219	65.32	1.55
2–3 part-time teachers	217	73.05	1.50
4–6 part-time teachers	181	72.21	1.63
> 7 part-time teachers	135	53.39	1.86
Full-time teachers certified			
0–90 percent of full-time teachers certified	197	63.95	1.62
91–99 percent of full-time teachers certified	135	70.97	1.65
100 percent of full-time teachers certified	420	69.04	1.60
School coeducational status			
Coeducational school	711	68.29	1.61
All-female school	19	42.42	1.64
All-male school	22	57.49	1.73
Total enrollment			
Total enrollment 0–600 students	189	70.37	1.52
Total enrollment 601–1,200 students	220	65.11	1.56
Total enrollment 1,201–1,800 students	171	63.81	1.57
Total enrollment >1,800 students	172	56.62	1.82
Census region			
Northeast	134	60.08	1.90
Midwest	189	73.81	1.45
South	281	70.34	1.37
West	148	62.72	2.02

See notes at end of table.

Table 33. Average weight adjustment factors used to adjust school weights for nonresponse: 2002–Continued

Model predictor variables[1]	Number of responding schools	Weighted response rate[2]	Average weight adjustment factor
Asian 10th-grade enrollment			
≤ 2 percent Asian 10th-grade enrollment	292	64.81	1.50
> 2 percent Asian 10th-grade enrollment	460	68.53	1.68
Black 10th-grade enrollment			
≤ 4 percent Black 10th-grade enrollment	255	61.73	1.78
> 4 percent Black 10th-grade enrollment	497	70.21	1.52
Hispanic 10th-grade enrollment			
≤ 3 percent Hispanic 10th-grade enrollment	289	69.23	1.56
> 3 percent Hispanic 10th-grade enrollment	463	66.99	1.64
Other 10th-grade enrollment			
≤ 80 percent Other 10th-grade enrollment	365	65.90	1.62
> 80 percent Other 10th-grade enrollment	387	68.67	1.60
CHAID[5] segments			
CHAID segment 1 = Northeast, 0–3 part-time teachers	87	61.19	1.62
CHAID segment 2 = Northeast, > 4 part-time teachers	665	68.23	1.61
CHAID segment 3 = Midwest and South, ≤ 4 percent Black 10th-grade enrollment	403	62.34	1.87
CHAID segment 4 = Midwest and South, > 4 percent Black 10th-grade enrollment, 1–40 full-time teachers	98	77.03	1.19
CHAID segment 5 = Midwest and South, > 4 percent Black 10th-grade enrollment, > 40 full-time teachers	251	72.90	1.36
CHAID segment 6 = West, ≤ 5 percent IEP	669	67.68	1.53
CHAID segment 7 = West, 6–10 percent IEP	39	41.83	2.68
CHAID segment 8 = West, > 10 percent IEP	44	77.81	1.87

[1]Model predictor variables had a value of 0 or 1. Some of the listed model predictor variables were not actually in the model because they served as reference groups. For each group of variables, one of the categories (predictor variable) was used as a reference group.
[2]Unrounded weights were used to calculate weighted response rates.
[3]IEP = Individual education program.
[4]LEP = Limited English proficient.
[5]CHAID = Chi-squared automatic interaction detection.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

GEMs were used to poststratify the school analysis weight. The counts of schools were controlled by school type, urbanicity, region, and enrollment, and control totals were formed based on the sampling frame. The sampling frame was used because there was not a more current PSS file available, and the CCD file that was available had some missing values for enrollment. However, the new schools that were added to the sample (see section 3.2.2) were not included in the poststratification adjustment. The outcome of this second model was a poststratification adjustment factor (WT5). Table 34 presents the control totals and the average weight adjustment factors needed to achieve these totals. The poststratification weight adjustment factors met the following constraints:

- minimum: 0.07
- median: 1.01
- maximum: 2.83

The final school weight used as the intermediate weight for the student weight was the product of WT1 – WT4, i.e. WT1*WT2*WT3*WT4. The final school weight used for school level analysis is the product of WT1 – WT5, that is WT1*WT2*WT3*WT4*WT5. Table 35 shows statistical properties of the school analysis weight. This school analysis weight is also applicable to school administrator data, library data, and facilities checklist data. Each of these had a response rate of greater than 95 percent, so one school weight will be sufficient.

Table 34. Average weight adjustment factors for poststratifying to control totals: 2002

Model variable[1]	Control total	Average weight adjustment factor
Total enrollment categories		1.05
10th–grade enrollment		
0–99 10th-grade students	16,841	1.29
100–249 10th-grade students	5,352	1.10
250–499 10th-grade students	3,777	0.97
≥ 500 10th-grade students	1,517	0.88
Urbanicity		
Urban	6,672	1.05
Suburban	11,857	1.02
Rural	8,958	1.15
Census region		
Northeast	4,262	1.02
Midwest	7,371	1.09
South	9,846	1.06
West	6,008	1.03
School type		
Public	20,408	1.03
Catholic	1,205	1.08
Other private	5,874	1.23
School type by enrollment category		
Public		
0–99 10th-grade students	10,581	1.35
100–249 10th-grade students	4,659	1.08
250–499 10th-grade students	3,659	0.98
≥ 500 10th-grade students	1,509	0.88
Catholic or other private		
0–99 10th-grade students	6,260	1.23
Catholic 100–249 10th-grade students	507	1.03
Catholic and ≥ 250 10th-grade students or other private and ≥ 100 10th-grade students	312	1.12
Census region by enrollment category		
Northeast		
0–249 10th-grade students	3,380	1.20
≥ 250 10th-grade students	882	0.82
Midwest		
0–99 10th-grade students	4,677	1.12
100–249 10th-grade students	1,524	1.27
≥ 250 10th-grade students	1,170	0.96
South		
0–99 10th-grade students	6,129	1.48
100–249 10th-grade students	1,829	0.92
250–499 10th-grade students	1,381	1.01
≥ 500 10th-grade students	507	0.91

See notes at end of table.

Table 34. Average weight adjustment factors for poststratifying to control totals: 2002–Continued

Model variable[1]	Control total	Average weight adjustment factor
West		
0–249 10th-grade students	4,654	1.23
250–499 10th-grade students	754	0.86
≥ 500 10th-grade students	600	1.00
Urbanicity by enrollment category		
Urban		
0-99 10th-grade students	3,907	1.39
100–249 10th-grade students	941	1.05
250–499 10th-grade students	1,194	1.12
≥ 500 10th-grade students	630	0.76
Suburban		
0–99 10th-grade students	6,138	1.15
100–249 10th-grade students	2,880	1.07
250–499 10th-grade students	2,080	0.93
≥ 500 10th-grade students	759	1.01
Rural		
0–99 10th-grade students	6,796	1.36
100–249 10th-grade students	1,531	1.30
≥ 250 10th-grade students	631	0.82
School type by urbanicity		
Public		
Urban	3,968	1.02
Suburban	8,392	1.01
Rural	8,048	1.07
Catholic		
Urban	635	1.11
Suburban or rural	570	1.04
Other private		
Urban	2,069	1.07
Suburban	2,939	1.10
Rural	866	2.16
School type by region		
Public		
Northeast	2,838	0.95
Midwest	5,908	1.02
South	7,088	1.05
West	4,574	1.04
Catholic		
Northeast, South, or West	807	0.97
Midwest	398	1.29
Other private		
Northeast, Midwest, or West	3,376	1.19
South	2,498	1.30

[1]Model predictor variables had a value of 0 or 1. Some of the listed model predictor variables were not actually in the model because they served as reference groups. For each group of variables, one of the categories (predictor variable) was used as a reference group.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 35. Statistical properties of school weight: 2002

Weight	BYSCHWT
Mean	32.97
Variance	1,185.67
Standard deviation	34.43
Coefficient of variation (X 100)	146.37
Minimum	1.00
Maximum	395.76
Skewness	3.61
Kurtosis	15.64
Sum	24,794.50
Number of cases	752

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.4.2 Student Weights

Two sets of student weights were computed. There is one set of weights for student questionnaire completion (BYWTSTU); this is the sole student weight that appears on 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 on the restricted-use file. This weight sums to the total of all 10th-grade students.[37]

First, the student-level design weight (WT6) 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 student stratum s within school I ($f_{s\backslash i}$)was calculated as the overall sampling rate divided by the school's probability of selection, or:

$$f_{s|i} = \frac{f_s}{\pi_r(i)},$$

where:

f_s = the overall student sampling rate, and

π_r (i) = the school's probability of selection.

If the school's enrollment list was larger than expected based on the CCD or PSS data, the preloaded student sampling rates yielded larger-than-expected sample sizes. Likewise, if the enrollment list was smaller than expected, the sampling rates yielded smaller-than-expected sample sizes. To maintain control of the sample sizes and to accommodate in-school data collection, the sampling rates were adjusted, when necessary, so that the number of students

[37] Students were excused from completion of the questionnaire or the test when (for reasons of their lack of proficiency in English or severe disabilities) they could not validly be assessed or complete a questionnaire, or could only be surveyed under conditions that would be unduly arduous or uncomfortable for them. It is anticipated that the expanded sample will be used primarily for methodological studies of the difference between the full and questionnaire/test-excluded populations and for calculation of a cohort dropout rate between 2002 and 2004 that takes account of all spring 2002 sophomores, regardless of their ability to complete the survey instruments. Students excluded from the instrument-completing portion of the sample will be re-evaluated in 2004 and will be surveyed if their language or disability status has changed such as to make fuller participation meaningful and possible.

selected did not exceed 35 students. A minimum sample size constraint of 10 students also was 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 rates, or:

$$WT6 = 1 / f_{s|i} .$$

The probability of selection for a refreshed student was equal to the probability of selection of the student that the refreshed student was linked to during selection using the half-open interval rule (see section 3.2.3.5).

When schools provided hardcopy lists for student sampling, they often did not provide separate lists by strata (e.g., students of all races were on the same list). When that happened, the combined list was sampled at the highest of the sampling rates for the strata contained within the list. After the original sample was keyed, strata with the lower sampling rates were then subsampled to achieve the desired sampling rates. The student subsampling weight (WT7) is the reciprocal of this subsampling rate. This weight is unity (1.00) for many students because this subsampling was not necessary for many schools.

Student eligibility was determined at the sampling stage, on Survey Day and Make-up Day(s), and during subsequent CATI follow-up. Eligibility was not determined for all nonrespondents. Attempts were made to contact all nonrespondents in CATI, when possible. Adjusting the weights of nonrespondents to compensate for the small portion of students who were actually ineligible was considered. However, in CATI, only nine ineligible students were identified, so it was assumed that all of the nonrespondents are eligible. If the assumption had been made that some nonrespondents were ineligible, the adjustment would be negligible.

For each set of student weights, adjustment factors were computed similarly but for a different population. GEMs were used, as described above for school weight adjustments. The variables available for most respondents and nonrespondents are described in section 3.2.6. For data known for most but not all students, data collected from responding students and weighted hot-deck imputation were used so that there were data for all eligible sample students.

The student nonresponse adjustment was performed in two stages—parent refusal and other nonresponse—because the predictors of response propensity were potentially different at each stage. The nonresponse models reduce the bias due to nonresponse for the model predictor variables and related variables. Therefore, using these two stages of nonresponse adjustment achieved greater reduction in nonresponse bias to the extent that different variables were significant predictors of response propensity at each stage. GEMs were used to compute the two student nonresponse adjustment factors (WT8 and WT9). Table 36 presents the final predictor variables used in the first stage student nonresponse adjustment model and the average weight adjustment factors resulting from these variables. The first stage of nonresponse adjustment factors met the following constraints:

- minimum: 0.10
- median: 1.08
- maximum: 2.25

Table 37 presents the final predictor variables used in the second-stage student nonresponse adjustment model and the average weight adjustment factors resulting from these variables. The second stage of nonresponse adjustment factors met the following constraints:

- minimum: 1.00
- median: 1.05
- maximum: 2.27

Table 36. Average weight adjustment factors used to adjust student weights for parent refusal: 2002

Model predictor variables[1]	Number of responding students and student refusals	Weighted response rate	Average weight adjustment factor
Total	16,309	93.05	1.11
School type			
Public schools	12,886	93.09	1.10
Catholic schools	1,958	92.29	1.09
Other private schools	1,465	92.96	1.21
Metropolitan status			
Urban	5,576	93.68	1.14
Suburban	7,773	92.26	1.10
Rural	2,960	94.14	1.09
10th-grade enrollment			
0–99 10th-grade students	3,146	95.67	1.12
100–249 10th-grade students	4,114	93.55	1.08
250–499 10th-grade students	5,272	92.81	1.11
> 500 10th-grade students	3,777	91.89	1.14
Type of grades within school			
K–12, PreK–10th, 1st–12th, PreK/1st–9th/12th and PreK–12 schools	1,035	94.71	1.27
Schools that contain middle grades but not elementary grades	1,719	95.50	1.06
Schools that only contain high school grades	13,555	92.75	1.10
Number of grades within the school			
4 grade levels within the school	1,111	93.69	1.09
> or < 4 grade levels within the school	15,198	93.00	1.11
Number of days in school year			
Less than 180 days of school in the school year	4,200	93.76	1.11
180 days of school in the school year	9,090	93.64	1.10
More than 180 days of school in the school year	3,019	90.36	1.12
Minutes per class period			
≤ 45 minutes per class period	3,928	92.39	1.12
46–50 minutes per class period	3,464	92.52	1.11
51–80 minutes per class period	4,331	92.22	1.14
≥ 81 minutes per class period	4,586	94.67	1.07
Class periods per day			
1–4 class periods per day	4,731	94.49	1.08
5–6 class periods per day	3,990	91.67	1.15
7 class periods per day	4,430	93.52	1.09
8–9 class periods per day	3,158	92.02	1.14

See notes at end of table.

Table 36. Average weight adjustment factors used to adjust student weights for parent refusal: 2002–Continued

Model predictor variables[1]	Number of responding students and student refusals	Weighted response rate	Average weight adjustment factor
IEP[3] percentage			
≤ 5 percent IEP	6,296	93.15	1.11
6–10 percent IEP	4,176	92.71	1.11
11–15 percent IEP	3,623	94.48	1.09
> 15 percent IEP	2,214	91.00	1.13
LEP[4] percentage			
0 percent LEP	6,972	94.24	1.10
1 percent LEP	3,208	91.65	1.10
2–5 percent LEP	2,831	94.03	1.08
> 6 percent LEP	3,298	91.96	1.16
Free or reduced price lunch			
0 percent free or reduced-price lunch	2,823	90.80	1.14
1–10 percent free or reduced-price lunch	3,672	90.77	1.11
11–30 percent free or reduced-price lunch	4,899	93.70	1.11
> 31 percent free or reduced-price lunch	4,915	94.75	1.09
Number of full-time teachers			
1–40 full-time teachers	4,193	95.40	1.08
41–70 full-time teachers	4,135	94.10	1.08
71–100 full-time teachers	4,230	91.91	1.14
> 100 full-time teachers	3,751	91.98	1.13
Number of part-time teachers			
0–1 part-time teachers	4,831	95.22	1.08
2–3 part-time teachers	4,623	92.68	1.11
4–6 part-time teachers	3,947	93.28	1.11
> 7 part-time teachers	2,908	89.80	1.14
Full-time teachers certified			
0–90 percent of full-time teachers certified	4,151	93.03	1.15
91–99 percent of full-time teachers certified	2,936	93.78	1.10
100 percent of full-time teachers certified	9,222	92.83	1.10
School coeducational status			
Coeducational school	15,507	93.08	1.11
All-female school	374	91.16	1.09
All-male school	428	91.89	1.09
Total enrollment			
Total enrollment 0–600 students	3,785	95.56	1.09
Total enrollment 601–1,200 students	4,906	93.24	1.10
Total enrollment 1,201–1,800 students	3,723	92.33	1.10
Total enrollment >1,800 students	3,895	92.06	1.15
Census region			
Northeast	3,044	91.08	1.13
Midwest	4,122	93.69	1.07
South	5,842	93.49	1.09
West	3,301	93.33	1.18

See notes at end of table.

Table 36. Average weight adjustment factors used to adjust student weights for parent refusal: 2002–Continued

Model predictor variables[1]	Number of responding students and student refusals	Weighted response rate	Average weight adjustment factor
Asian 10th-grade enrollment			
≤ 2 percent Asian 10th-grade enrollment	6,327	93.35	1.08
> 2 percent Asian 10th-grade enrollment	9,982	92.87	1.13
Black 10th-grade enrollment			
≤ 4 percent Black 10th-grade enrollment	5,486	92.29	1.12
> 4 percent Black 10th-grade enrollment	10,823	93.45	1.10
Hispanic 10th-grade enrollment			
≤ 3 percent Hispanic 10th-grade enrollment	6,263	92.36	1.09
> 3 percent Hispanic 10th-grade enrollment	10,046	93.48	1.12
Other 10th-grade enrollment			
≤ 80 percent Other 10th-grade enrollment	8,196	93.71	1.12
> 80 percent Other 10th-grade enrollment	8,113	92.36	1.10
CHAID[4] segments			
CHAID segment 1 = 1–40 full-time teachers, public school, ≤ 2 percent Asian 10th-grade enrollment	372	91.35	1.09
CHAID segment 2 = 1–40 full-time teachers, public school, > 2 percent Asian 10th-grade enrollment	15,937	93.09	1.11
CHAID segment 3 = 1–40 full-time teachers, Catholic and other private schools, Hispanic and Other race	16,050	93.04	1.11
CHAID segment 4 = 1-40 full-time teachers, Catholic and other private schools, Asian and Black	259	97.54	1.13
CHAID segment 5 = 41–70 full-time teachers, 0–6 part-time teachers, 1–6 class periods	1,894	96.30	1.07
CHAID segment 6 = 41–70 full-time teachers, 0–6 part-time teachers, 7–9 class periods	14,415	92.65	1.11
CHAID segment 7 = 41–70 full-time teachers, > 7 part-time teachers, ≤ 180 school days	703	92.09	1.11
CHAID segment 8 = 41–70 full-time teachers, > 7 part-time teachers, > 180 school days	15,606	93.09	1.11
CHAID segment 9 = > 70 full-time teachers, 0–1 part time teachers, ≤ 80 percent Other 10th-grade enrollment	15,667	93.17	1.11
CHAID segment 10 = > 70 full-time teachers, 0–1 part time teachers, >80 percent other 10th-grade enrollment	642	90.91	1.15
CHAID segment 11 = > 70 full-time teachers, >=2 part-time teachers, ≤ 45 minutes per class	791	88.20	1.20
CHAID segment 12 = > 70 full-time teachers, ≥ 2 part-time teachers, 46–80 minutes per class	13,782	93.18	1.11
CHAID segment 13 = > 70 full-time teachers, ≥ 2 part-time teachers, ≥ 81 minutes per class	1,736	94.11	1.08
Student sex			
Male	8,203	92.84	1.12
Female	8,106	93.27	1.10

See notes at end of table.

Table 36. Average weight adjustment factors used to adjust student weights for parent refusal: 2002–Continued

Model predictor variables[1]	Number of responding students and student refusals	Weighted response rate	Average weight adjustment factor
Student race/ethnicity			
Hispanic	2,061	94.07	1.08
Other	10,022	92.83	1.11
Black	2,471	94.08	1.09
Asian	1,755	89.05	1.16

[1]Model predictor variables had a value of 0 or 1. Some of the listed model predictor variables were not actually in the model because they served as reference groups. For each group of variables, one of the categories (predictor variable) was used as a reference group.
[2]IEP = Individualized education program.
[3]LEP = Limited English proficient.
[4]CHAID = Chi-squared automatic interaction detection.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 37. Average weight adjustment factors used to adjust student weights for other nonresponse: 2002

Model predictor variables[1]	Number of respondin g students	Weighted response rate	Average weight adjustment factor
Total	15,362	93.71	1.06
School type			
Public schools	12,039	93.44	1.07
Catholic schools	1,920	97.76	1.02
Other private schools	1,403	96.08	1.04
Metropolitan status			
Urban	5,115	90.47	1.09
Suburban	7,399	94.76	1.05
Rural	2,848	96.04	1.04
10th-grade enrollment			
0–99 10th-grade students	3,041	96.95	1.04
100–249 10th-grade students	3,976	96.70	1.03
250–499 10th-grade students	4,941	93.32	1.07
> 500 10th-grade students	3,404	90.65	1.11
Type of grades within school			
K–12, PreK–10th, 1st–12th, PreK/1st–9th/12th and PreK–12 schools	998	97.38	1.03
Schools that contain middle grades but not elementary grades	1,647	95.87	1.04
Schools that only contain high school grades	12,717	93.33	1.07
Number of grades within the school			
4 grade levels within the school	1,027	92.76	1.08
> or < 4 grade levels within the school	14,335	93.79	1.06
Number of days in school year			
Less than 180 days of school in the school year	4,019	95.20	1.04
180 days of school in the school year	8,522	93.40	1.07
More than 180 days of school in the school year	2,821	92.73	1.07
Minutes per class period			
≤ 45 minutes per class period	3,655	91.33	1.08
46–50 minutes per class period	3,295	95.00	1.05
51–80 minutes per class period	4,063	93.51	1.07
≥ 81 minutes per class period	4,349	94.51	1.06
Class periods per day			
1–4 class periods per day	4,473	94.39	1.06
5–6 class periods per day	3,715	92.63	1.08
7 class periods per day	4,178	93.71	1.06
8–9 class periods per day	2,996	94.24	1.06
IEP[3] percentage			
≤ 5 percent IEP	5,961	93.25	1.06
6–10 percent IEP	3,908	93.27	1.07
11–15 percent IEP	3,442	95.10	1.05
> 15 percent IEP	2,051	93.02	1.08
LEP[4] percentage			
0 percent LEP	6,701	95.77	1.04
1 percent LEP	3,041	94.61	1.05
2–5 percent LEP	2,579	91.48	1.10
> 6 percent LEP	3,041	91.72	1.09

See notes at end of table.

Table 37. Average weight adjustment factors used to adjust student weights for other nonresponse: 2002–Continued

Model predictor variables[1]	Number of responding students	Weighted response rate	Average weight adjustment factor
Free or reduced price lunch			
0 percent free or reduced-price lunch	2,722	96.03	1.03
1–10 percent free or reduced-price lunch	3,459	93.46	1.06
11–30 percent free or reduced-price lunch	4,631	94.45	1.06
> 31 percent free or reduced-price lunch	4,550	92.48	1.08
Number of full-time teachers			
1–40 full-time teachers	4,066	96.86	1.04
41–70 full-time teachers	3,927	94.77	1.05
71–100 full-time teachers	3,921	92.73	1.08
> 100 full-time teachers	3,448	91.96	1.09
Number of part-time teachers			
0–1 part-time teachers	4,470	92.23	1.08
2–3 part-time teachers	4,432	95.26	1.05
4–6 part-time teachers	3,749	95.04	1.05
> 7 part-time teachers	2,711	92.34	1.07
Full-time teachers certified			
0–90 percent of full-time teachers certified	3,898	92.51	1.06
91–99 percent of full-time teachers certified	2,743	93.00	1.07
100 percent of full-time teachers certified	8,721	94.25	1.06
School coeducational status			
Coeducational school	14,577	93.63	1.06
All-female school	365	97.35	1.03
All-male school	420	98.26	1.02
Total enrollment			
Total enrollment 0–600 students	3,685	97.27	1.03
Total enrollment 601–1,200 students	4,655	94.86	1.05
Total enrollment 1,201–1,800 students	3,492	93.43	1.07
Total enrollment >1,800 students	3,530	90.81	1.10
Census region			
Northeast	2,763	90.12	1.10
Midwest	3,897	93.60	1.06
South	5,622	96.11	1.04
West	3,080	93.16	1.07
Asian 10th-grade enrollment			
≤ 2 percent Asian 10th-grade enrollment	5,963	93.90	1.06
> 2 percent Asian 10th-grade enrollment	9,399	93.60	1.06
Black 10th-grade enrollment			
≤ 4 percent Black 10th-grade enrollment	5,214	94.51	1.05
> 4 percent Black 10th-grade enrollment	10,148	93.30	1.07
Hispanic 10th-grade enrollment			
≤ 3 percent Hispanic 10th-grade enrollment	5,974	95.09	1.05
> 3 percent Hispanic 10th-grade enrollment	9,388	92.88	1.07
Other 10th-grade enrollment			
≤ 80 percent Other 10th-grade enrollment	7,582	92.08	1.08
> 80 percent Other 10th -grade enrollment	7,780	95.43	1.04

See notes at end of table.

Table 37. Average weight adjustment factors used to adjust student weights for other nonresponse: 2002–Continued

Model predictor variables[1]	Number of responding students	Weighted response rate	Average weight adjustment factor
CHAID[4] segments			
CHAID segment 1 = Race = Hispanic, Asian, Black, Northeast, 0-499 10th-grade students	15,079	94.30	1.06
CHAID segment 2 = Race = Hispanic, Asian, Black, Northeast, ≥ 500 10th-grade students	283	67.34	1.36
CHAID segment 3 = Race = Hispanic, Asian, Black, Midwest, 0–90 percent full-time teachers certified	184	76.23	1.24
CHAID segment 4= Race = Hispanic, Asian, Black, Midwest, 91-100 percent full-time teachers certified	15,178	93.86	1.06
CHAID segment 5 = Race = Hispanic, Asian, Black, South, 0–99 10th-grade students	178	97.64	1.02
CHAID segment 6 = Race = Hispanic, Asian, Black, South, 100–499 10th-grade students	14,551	93.71	1.06
CHAID segment 7 = Race = Hispanic, Asian, Black, South, ≥ 500 10th-grade students	633	93.07	1.07
CHAID segment 8 = Race= Hispanic, Asian, Black, West, 0–499 10th-grade students	770	96.26	1.07
CHAID segment 9 = Race= Hispanic, Asian, Black, West, ≥ 500 10th-grade students	14,592	93.64	1.06
CHAID segment 10 = Race = Other, 0–249 10th-grade students, 1–40 full-time teachers	3,125	97.71	1.02
CHAID segment 11 = Race = Other, 0–249 10th-grade students, 41–100 full-time teachers	12,168	93.06	1.07
CHAID segment 12 = Race = Other, 0–249 10th-grade students, > 100 full-time teachers	69	91.61	1.04
CHAID segment 3 = Race = Other, ≥ 250 10th-grade students, Northeast and Midwest	1,786	92.09	1.08
CHAID segment 14 = Race = Other, ≥ 250 10th-grade students, South and West	13,576	94.04	1.06
Student sex			
Male	7,658	92.93	1.07
Female	7,704	94.52	1.05
Student race/ethnicity			
Hispanic	1,894	90.55	1.09
Asian	9,585	95.08	1.04
Black	2,257	91.07	1.09
Other	1,626	92.12	1.09

[1]Model predictor variables had a value of 0 or 1. Some of the listed model predictor variables were not actually in the model because they served as reference groups. For each group of variables, one of the categories (predictor variable) was used as a reference group.
[2] IEP = Individualized education plan.
[3] LEP = Limited English proficiency.
[4] CHAID = Chi-squared automatic interaction detection.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Poststratification was not necessary for students because the collected student data were more current than any outside sources. However, the student counts were compared to the frame counts by region, urbanicity, and school type, and after accounting for ineligibles on the frame, the counts seemed reasonable.

Both of the final student weights are the product of the school weight and the appropriate WT6 – WT9, i.e., (FINAL SCHOOL WEIGHT)*WT6*WT7*WT8*WT9. Table 38 shows statistical properties of the student weights.

Table 38. Statistical properties of student weights: 2002

Weight	BYSTUWT
Mean	223.90
Variance	18,597.52
Standard deviation	136.37
Coefficient of variation (X 100)	67.02
Minimum	5.09
Maximum	978.38
Skewness	0.99
Kurtosis	0.99
Sum	3,439,489.61
Number of cases	15,362

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

3.4.3 Quality Control

Quality control was emphasized on all activities, including weighting. Because of the central importance of the analysis weights to population estimation, a senior statistician also thoroughly checked each set of weights. The most fundamental type of check was the verification of totals that are algebraically equivalent (e.g., marginal totals of the weights of eligible schools or students prior to nonresponse adjustment and of respondents after nonresponse adjustment). In addition, various analytic properties of the initial weights, the weight adjustment factors, and the final weights were examined both overall and within sampling strata, including:

- distribution of the weights;
- ratio of the maximum weight divided by the minimum weight; and
- unequal weighting design effect, or variance inflation effect ($1 + CV^2$).

Additionally, two-dimensional tables were reviewed before and after weight adjustments to ensure that the weight distribution was not distorted.

3.5 Standard Errors and Design Effects

3.5.1 Standard Errors

For probability-based sample surveys, most estimates are nonlinear statistics. For example, a mean or proportion, which is expressed as $\Sigma wy/\Sigma w$,[38] is nonlinear because the denominator is a survey estimate of the (unknown) population total. In this situation, the variances of the estimates cannot be expressed in closed form. 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. Woodruff presented the mathematical formulation of this procedure.[39] The variance estimation must also take into account stratification and clustering. There are other variance estimation procedures, such as jackknife and balanced repeated replication (BRR). However, Taylor series estimation was determined to be sufficient for ELS:2002.

For stratified multistage surveys, the Taylor series procedure requires analysis strata and analysis PSUs. School sampling strata exist and the PSUs are individual schools. However, given that the school sample was selected using probability with minimum replacement (pmr), it is recommended for variance estimation that there are two PSUs per stratum.[40] Therefore, analysis strata were defined from the sampling strata used in the first stage of sampling. The responding schools were sorted within sampling strata in the same order as was used for sampling, and then adjacent analysis PSUs were paired to form analysis strata. When there was an odd number of schools in a sampling stratum, then one of the analysis strata formed had three PSUs. This procedure resulted in 361 analysis strata.

As described in section 3.2, the ELS:2002 sampling design was a stratified two-stage design. A stratified sample of schools was selected with probabilities proportional to a composite measure of size at the first stage, and a stratified systematic sample of students was selected from sample schools at the second stage. At the first stage, the school sampling rates varied considerably by school sampling strata. At the second stage, Asian and Hispanic students were sampled at higher rates than other students. Because of this complex sampling design, statistical analyses should be conducted using software that properly accounts for the complex survey design.

Many commonly used statistical computing packages assume that the data were obtained from a simple random sample; that is, they assume that the observations are independent and identically distributed. When the data have been collected using a complex sampling design, the simple random sampling assumption usually leads to an underestimate of the sampling variance, which would lead to artificially small confidence intervals and liberal hypothesis test results (i.e., rejecting the null hypothesis when it is in fact true more often than indicated by the nominal Type I error level).[41]

[38] w is the estimated population, and y is a 0/1 variable indicating whether or not a certain characteristic is present for the sample member.
[39] See Woodruff (1971).
[40] See Chromy (1981).
[41] See Carlson, Johnson, and Cohen (1993).

Statistical strategies that have been developed to address this issue include first-order Taylor series expansion of the variance equation, balanced repeated replication, and the Jackknife approach.[42] Special-purpose software packages that have been developed for analysis of complex sample survey data include SUDAAN, WesVar, and Stata. Evaluations of the relative performances of these packages are reported by Cohen.[43] SUDAAN is a commercial product developed by RTI; information regarding the features of this package and its lease terms is available from the web site *http://www.rti.org/sudaan*. WesVar is a product of Westat, Inc.; information regarding the features of this package and its lease terms is available from the web site *http://www.westat.com/wesvar*. In addition to the variance estimation packages noted above, the National Center for Education Statistics has recently co-sponsored the development of the AM variance estimation software. AM software can be downloaded for free from the following web site: *http://am.air.org/*.

Below is an example of generic SUDAAN code to produce estimates and standard errors using Taylor Series. The symbols /* and */ in the code indicate the beginning and end of a comment. Note that the data set must be sorted by analysis strata and analysis PSUs.

```
proc descript data=/* insert filename*/ design=wr;
nest analstr analpsu; /* these variables are the analysis strata and analysis PSUs,
respectively */
weight BYQWTSTU;
var /*insert variables*/;
subpopn /* insert domain of interest if domain is a subset of students*/;
print nsum mean semean  / style=nchs;
run;
```

3.5.2 Design Effects

The impact of the departures 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.[44] The design effect is the ratio of the actual variance of the statistic to the variance that would have been obtained had the sample been a simple random sample. The design standard errors will be different from the standard errors that are based on the assumption that the data are from a simple random sample. The ELS:2002 sample departs from the assumption of simple random sampling in three major respects: both schools and student samples were stratified by school and student characteristics, respectively; both schools and 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.

[42] See Wolter (1985).
[43] See Cohen (1997).
[44] The variance due to imputation was not taken into account in the computation of the design effect.

Standard errors and design effects were computed at the first stage (school level) and at the second stage (student level). There are multiple instruments at both levels. At the school level, there was a school administrator questionnaire, a library media center questionnaire, and a facilities checklist. The school administrator questionnaire was the basis for the school-level calculations, however, two items from the library questionnaire were also included. At the school level, there were student, parent, and teacher questionnaires. 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), and this 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 school subgroups are similar to those used in NELS:88:

- school type (public and all private); and
- school 10th-grade enrollment (large versus small).[45]

The student and parent subgroups are also similar to those used in NELS:88:

- sex (male and female);
- race/ethnicity (Asian or Pacific Islander, American Indian or Alaska native, Black, Hispanic, White or other, Multiracial);
- school type (public, Catholic, and other private);
- socioeconomic status (SES) (lowest quartile, middle two quartiles, and highest quartile); and
- urbanicity (urban, suburban, and rural).

Critical school items were identified as the 12 items included in both the school characteristics questionnaire for nonresponding schools and the school administrator questionnaire. Sixteen additional school variables were selected randomly from the administrator questionnaire, and two variables were selected randomly from the library questionnaire.

Four variables that are row variables in most of the six chapters of the *Profile of the American High School Sophomore in 2002* report were selected. Additionally three items were randomly chosen from each of four chapters and two items were randomly chosen from each of two chapters for a total of 16 items. This approach guarantees a range of data that will give a reasonable average, as well as a reading on design effects for subgroups. Also, most of the trend variables are included in this report, which will maximize comparability of design effect results with HS&B and NELS:88. Finally, 10 student items were chosen specifically because they were used in NELS:88.

Nine critical parent items were identified, and 21 additional items were selected randomly from the parent questionnaire.

[45] Large schools are defined as those with 10th-grade enrollment of at least 300, and small schools are defined as those with 10th-grade enrollment fewer than 300.

The student variables used were the values after imputation and all variables used were after disclosure avoidance (see sections 3.3 and 3.6). Also, the public versions of the variables were used when the public version differed from the restricted version.

Appendix K contains tables of school, student, and parent variables. Each table includes the survey item (or composite variable), the variable name, percent estimate, design standard error, simple random sample standard error, sample size (N), the design effect (DEFF), and the square root of the design effect (DEFT). Tables 39, 40, and 41 summarize the average DEFFs and DEFTs for schools, students, and parents, respectively, for each subgroup.

Table 39. Mean design effects (DEFFs) and root design effects (DEFTs) for school and library questionnaire data: 2002

Group	Mean DEFF	Mean DEFT
All schools	2.76	1.64
Public schools	2.86	1.65
All private schools	2.63	1.59
Large schools[1]	2.07	1.43
Small schools[2]	1.04	1.01

[1]Large schools are defined as those with 10th-grade enrollment of at least 300.
[2]Small schools are defined as those with 10th-grade enrollment fewer than 300.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 40. Mean design effects (DEFFs) and root design effects (DEFTs) for student questionnaire data: 2002

Group	Mean DEFF	Mean DEFT
All students	2.35	1.50
Male	1.90	1.37
Female	2.01	1.40
White and other, non-Hispanic	2.03	1.41
Black, non-Hispanic	1.67	1.28
Hispanic or Latino	1.82	1.32
Asian or Pacific Islander	2.27	1.49
American Indian or Alaska native	1.42	1.18
Multiracial, non-Hispanic	1.63	1.27
Public schools	2.07	1.41
Catholic schools	2.43	1.51
Other private schools	3.53	1.78
Low socioeconomic status	1.70	1.29
Middle socioeconomic status	1.73	1.31
High socioeconomic status	1.99	1.39
Urban	2.88	1.64
Suburban	2.15	1.44
Rural	1.94	1.37

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

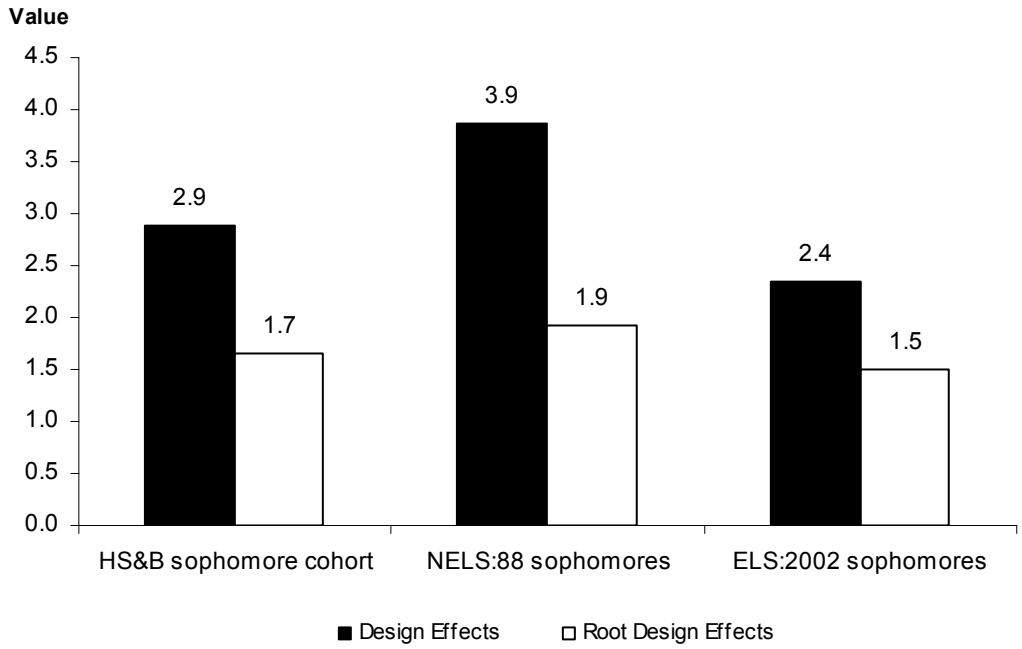
Table 41. Mean design effects (DEFFs) and root design effects (DEFTs) for parent questionnaire data: 2002

Group	Mean DEFF	Mean DEFT
All students	2.24	1.47
Male	1.79	1.33
Female	1.92	1.37
White and other, non-Hispanic	1.99	1.38
Black, non-Hispanic	1.48	1.21
Hispanic or Latino	1.66	1.28
Asian or Pacific Islander	1.97	1.39
American Indian or Alaska native	1.50	1.21
Multiracial, non-Hispanic	1.63	1.27
Public schools	1.98	1.38
Catholic schools	1.91	1.34
Other private schools	2.66	1.57
Low socioeconomic status	1.68	1.27
Middle socioeconomic status	1.61	1.26
High socioeconomic status	1.91	1.37
Urban	2.57	1.57
Suburban	2.13	1.44
Rural	1.86	1.31

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

The student-level design effects indicate that the ELS:2002 sample was more efficient than the NELS:88 sample and the High School and Beyond (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. This difference is also apparent for some subgroup estimates. Ingels et al. (1994) present design effects for 16 subgroups defined similarly to those in table 40 above. For all 16 subgroups, the ELS:2002 design effects are smaller on the average than those for the NELS:88 sophomores. Frankel et al. (1981) also present design effects for eight subgroups defined similarly to those in table 40 above. For all eight subgroups, the ELS:2002 design effects are smaller on average than those for the HS&B sophomore cohort. Figure 5 shows the mean design effects and root design effects for HS&B sophomores, NELS:88 sophomores, and ELS:2002 sophomores.

Figure 5. HS&B, NELS:88, and ELS:2002 mean design effects and root design effects: 2002



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

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. Although the general tendency in longitudinal studies is for design effects to lessen over time, as dispersion reduces the original clustering, subsampling increases design effects. This is because subsampling introduces additional variability into the weights with an attendant loss in sample efficiency.

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 HS&B base year sample design called for 36 sophomores selected from each school. The ELS:2002 sample design called for about 26 sophomores selected from each school. Clustering tends to increase the variance of survey estimates because the observations within a cluster are similar and therefore add less information than independently selected observations. The impact of clustering depends mainly on two factors: the number of observations within each cluster and the degree of within-cluster homogeneity. When cluster sizes vary, the impact of clustering (DEFFc) can be estimated by:

$$\text{DEFFc} = 1 + \left(\bar{b} - 1\right)\text{rho},$$

where $\bar{b}$ refers to the average cluster size (the average number of students selected from each school) and rho refers to the intraclass correlation coefficient, a measure of the degree of within-cluster homogeneity. If the value of rho (which varies from one variable to the next) averaged

about 0.05 in both studies, then the reduced cluster size in ELS:2002 would almost exactly account for the reduction in the design effects relative to HS&B.

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.

For all but two subgroups (American Indian or Alaska native and Multiracial, non-Hispanic), the average design effect for student items is larger than the average design effect for parent items. This suggests that the homogeneity of student responses within clusters is greater than the homogeneity of parent responses within the domains. Given the students' generally shared school experiences and the generally uniform questionnaire administration procedures this outcome is not surprising.

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 (see section 3.2.2). This is different from the methodology used for NELS:88 (see Spencer et al. 1991). The NELS:88 average school design effect (in the base year study) was considerably lower, 1.82.

If one must perform a quick analysis of ELS:2002 data without using one of the software packages for analysis of complex survey data, the design effects tables in appendix K can be used to make approximate adjustments to the standard errors of survey statistics computed using the standard software packages that assume simple random sampling designs. One cannot be confident regarding the actual design-based standard error without performing the analysis using one of the software packages specifically designed for analysis of data from complex sample surveys.

Standard errors for a proportion can be estimated from the standard error computed using the formula for the standard error of a proportion based on a simple random sample and the appropriate DEFT:

$$SE = DEFT * (p(1-p)/n)^{1/2}.$$

Similarly, the standard error of a mean can be estimated from the weighted variance of the individual scores and the appropriate mean DEFT:

$$SE = DEFT * (Var/n)^{1/2}.$$

Tables 39, 40, and 41 make it clear that the DEFFs and DEFTs vary considerably by subgroup. It is therefore important to use the mean DEFT for the relevant subgroup in calculating approximate standard errors for subgroup statistics.

Standard error estimates may be needed for subgroups that are not shown in the appendix. One rule of thumb may be useful in such situations. The general rule states that design effects will generally be smaller for groups that are formed by subdividing the subgroups listed in the tables. (This is because smaller subgroups will be affected less by clustering than larger subgroups; in terms of the equation for DEFFc, $\bar{b}$ will be reduced.) Estimates for Hispanic males, for example, will generally have smaller design effects than the corresponding estimates for all Hispanics or all males. For this reason, it will usually be conservative to use the subgroup mean DEFT to approximate standard errors for estimates concerning a portion of the subgroup. This rule only applies when the variable used to subdivide a subgroup crosscuts schools. Sex is one such variable because most schools include students of both sexes. It will not reduce the average cluster size to form groups that are based on subsets of schools.

Standard errors may also be needed for other types of estimates than the simple means and proportions that are the basis for the results presented in the above tables. A second method of procedure can be used to estimate approximate standard errors for comparisons between subgroups. If the subgroups crosscut schools, then the design effect for the difference between the subgroup means will be somewhat smaller than the design effect for the individual means; consequently, the variance of the difference estimate will be less than the sum of the variances of the two subgroup means from which it is derived:

$$Var(b\text{-}a) = Var(b) + Var(a)$$

where Var(b-a) refers to the variance of the estimated difference between the subgroup means, and Var(a) and Var(b) refer to the variances of the two subgroup means. This equation assumes that the covariance of the subgroup means is negligible. It follows from this equation that Var(a) + Var(b) can be used in place of Var(b-a) with conservative results.

A final principle is that more complex estimators show smaller design effects than simple estimators (Kish and Frankel 1974). Thus, correlation and regression coefficients tend to have smaller design effects than subgroup comparisons, and subgroup comparisons have smaller design effects than means. This implies that it will be conservative to use the DEFTs in the above tables in calculating approximate standard errors for complex statistics, such as multiple regression coefficients. The procedure for calculating such approximate standard errors is the same as with simpler estimates: first, a standard error is calculated using the formula for data from a simple random sample; then the standard error is multiplied by the appropriate DEFT.

One analytic strategy for accommodating complex survey designs is to use the mean design effect to adjust for the effective sample size resulting from the design. For example, one could create a weight that is the multiplicative inverse of the design effect and use that weight (in conjunction with sampling weights) to deflate the obtained sample size to take into account the inefficiencies due to a sample design that is a departure from a simple random sample. Using this procedure, statistics calculated by a statistical program such as SAS or SPSS will reflect the reduction in sample size in the calculation of standard errors and degrees of freedom. Such techniques capture the effect of the sample design on sample statistics only approximately. However, while not providing a full accounting of the sample design, this procedure provides some adjustment for the sample design, and is probably better than conducting analysis that assumes the data were collected from a simple random sample. The analyst applying this

correction procedure should carefully examine the statistical software being used, and assess whether or not the program treats weights in such a way as to produce the effect described above.

3.6 Disclosure Risk Analysis and Protections

Because of the paramount importance of protecting the confidentiality of NCES data that contain information about specific individuals, ELS:2002 data were subject to various procedures to minimize disclosure risk.

As a first step, all ELS:2002 data files (school, student, teacher, and parent) were reviewed to identify high risk variables. Some variables were identified as unsuitable for the public-use file in any form; these variables appear only on the restricted-use files. Public-use variables that might point to specific individuals or schools (e.g., some fine-grained variables, particularly those in continuous form, and variables with extreme outliers) were altered through data coarsening techniques, such as top coding, bottom coding, or recasting into categorical form.

As a second step, a technique called "data swapping" was carried out, both for school-level data, and for student-level data (student, parent, and teacher). Schools and respondents were randomly selected for swapping to achieve a specific, but undisclosed, swapping rate. In data swapping, some variables for a sample case that has been paired with another case will be exchanged. By so doing, even if a tentative identification of an individual is made, because every case in the file has some undisclosed probability of having been swapped, uncertainty remains about the accuracy and interpretation of the match.

As a final step, the ELS:2002 data underwent a disclosure risk analysis. In this analysis, school characteristics information available on the data files was compared to information on publicly available universe files of schools. A distance measure was used to compute risk of deductive disclosure, and techniques to minimize disclosure risk were applied until school identities were appropriately masked. Specific techniques employed included both perturbation (perturbation directly alters individual respondent data for some variables) and coarsening of the data (coarsening reduces the level of detail, for example, by making a continuous variable categorical).[46]

In the case of the coarsening applied to certain variables on the public-use file, more fine-grained detail for these variables may be found on the restricted-use files. In the case of perturbation of the data (including swapping), all changes imposed on the public-use files were also implemented in the restricted-use files. While perturbation techniques such as swapping do result in changes in estimates generated from the data, before-and-after weighted distributions and correlations for swapped variables show that after applying the disclosure limitation techniques, the analytic utility of the data files has not been compromised in any way.

[46] The NCES Statistical Standards (*http://nces.ed.gov/statprog/2002/std4_2.asp*), specifically NCES Standard 4-2, provide information both about the legislative background and legal requirements of maintaining confidentiality, and definitions of key terms (perturbation, coarsening, disclosure risk analysis, data swapping, and so on).

Chapter 4
Data Collection Methodology and Results

4.1 Data Collection Overview

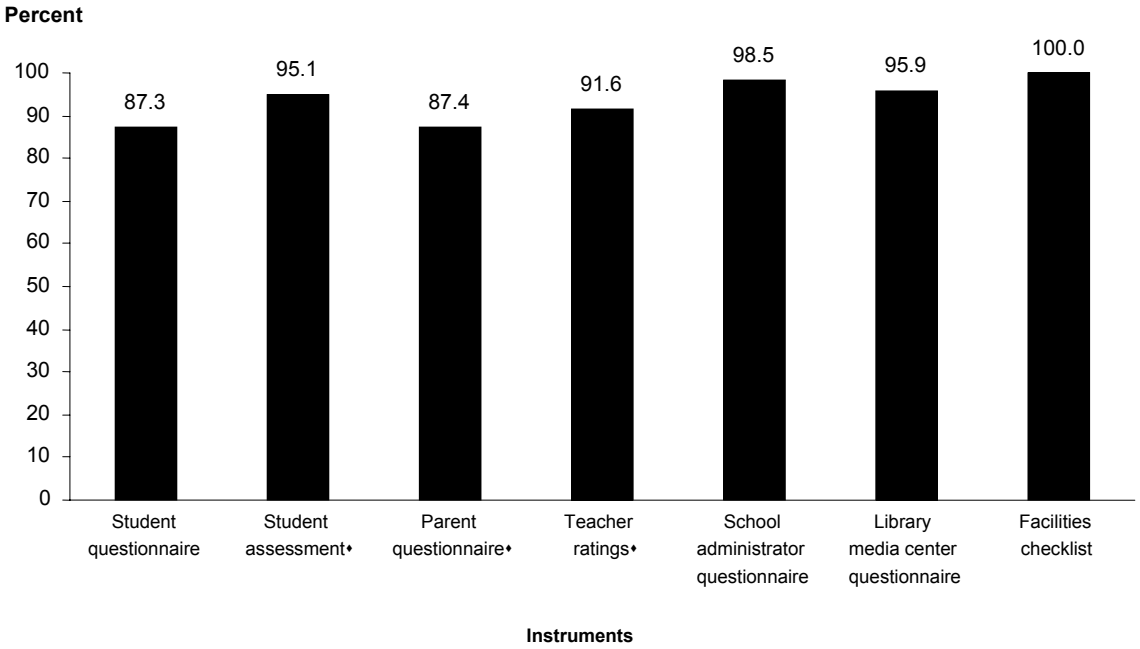
This chapter describes the data collection procedures for students for the base year of the Education Longitudinal Study of 2002 (ELS:2002). Data collection procedures for sources of contextual data (i.e., school administrators, librarians, teachers, parents, and facilities) are also discussed. Student data collection began in schools on January 21, 2002, and ended in the 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. Parent data collection ended on October 17, 2002. Results are summarized in table 42 and in figure 6 and provided in detail later in the chapter.

Table 42. Summary of ELS:2002 base year completion and coverage rates: 2002

Instrument	Selected	Participated	Weighted percent	Unweighted percent
Student questionnaire	17,591	15,362	87.28	87.33
Student assessment[1]	15,362	14,543	95.08	94.67
Parent questionnaire[2]	15,362	13,488	87.45	87.80
Teacher ratings of students[3]	15,362	14,081	91.64	91.66
School administrator questionnaire	752	743	98.53	98.80
Library media center questionnaire	752	718	95.93	95.48
Facilities checklist	752	752	100.00	100.00

[1]Percentage of cases for which a student questionnaire was obtained for which a cognitive test was also obtained. When a test was not obtained, test results were imputed.
[2]Indicates a coverage rate, the proportion of participating students with a parent report. More parents participated; completed case numbers reflect the records in the public-use data file, where parent (and teacher) data were excluded for students who did not complete a base year student questionnaire.
[3]Indicates a coverage rate: ratings obtained from at least one teacher.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Figure 6. Completion and coverage rates for ELS:2002 base year: 2002



♦Denotes a coverage rate.
NOTE: All completion rates are weighted.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

4.2 Study Endorsements, School Recruitment, and Pre-data Collection Activities

4.2.1 Study Endorsements

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 in the study (school administrators, librarians, teachers, students, and parents). The following organizations provided endorsements:

American Association of School Administrators
American Association of School Librarians
American Federation of Teachers
Council of Chief State School Officers
Council of the Great City Schools
National Association of Independent Schools
National Association of Secondary School Principals
National Catholic Educational Association, Department of Secondary Schools
National Education Association
National PTA
National Resource Center for Safe Schools
National School Boards Association

National School Safety Center
Seventh-Day Adventist Church

The list of endorsing organizations was included on the ELS:2002 letterhead for the lead letters that were sent at the state, district, and school levels. Endorsing agencies were also listed in an informational brochure and on the ELS:2002 public web site. Some organizations gave additional help in refusal conversion efforts with principals or headmasters.

4.2.2 School Recruitment

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 begin the approval process. Each CSSO was sent an information package. The package was addressed to the CSSO and contained a lead letter from the National Center for Education Statistics (NCES) project officer, a letter from the RTI project director, a study brochure, and a sample endorsement letter. The packages were sent by Federal Express so that it would be possible to track receipt.

About 1 week after sending the information package, the CSSOs were contacted by telephone. Project staff contacted CSSOs in states with particularly large numbers of schools or states in which there was reason to believe that the environment might make cooperation less likely; the staff of institutional recruiters contacted the remaining CSSOs. At that time, it was confirmed that the package had been received and it was determined who had been given responsibility for approving the study for the state. That person was then contacted to answer any questions and discuss participation. When asked, the state officials were provided with the number of schools and districts selected from their state, but for reasons of confidentiality, no districts or schools were named.

Permission to proceed to the district level was obtained in all 50 states as well as the District of Columbia. At the time permission was obtained at the state level, a person at the state level was identified to serve as a point of contact to address any questions from the districts about the state's participation. States were asked to provide a letter of endorsement from the state. A sample letter was provided as a template that the states could follow, if desired. A postage-paid envelope addressed to RTI was included to facilitate return of an endorsement letter. Endorsement letters were received from 40 states and the District of Columbia.

Once state approval was obtained, an information package was sent to each district/diocese that had sampled schools in the state. The package was addressed to the superintendent and sent by Federal Express. The package contained a lead letter from the NCES project officer, a letter from the RTI project director, a study brochure, a list of endorsing agencies, the state endorsement letter (if provided), and a sample endorsement letter.

Several days after sending the information package, the superintendents were contacted by telephone. The staff of institutional recruiters conducted telephone contacts with the districts. At the time of the call, it was confirmed that the package had been received and it was determined who had been given responsibility for approving the study for the district/diocese. That person was then contacted to answer any questions and discuss participation.

Permission to proceed to the school level was received from 693 of the 829 districts/ dioceses having eligible sampled schools (83.6 percent).[47] This represented a total of 891 eligible schools that had district/diocese permission to contact, among 1,059 eligible schools affiliated with districts/dioceses (84.1 percent).[48] As at the state level, approving districts/ dioceses were asked to identify a contact person at that level and to send a letter of endorsement. Endorsement letters were received from 148 districts/dioceses.

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 approvals, though endorsements from various private school organizations were sought.

As at the state and district levels, each school was sent an informational package by Federal Express. The package was addressed to the principal and contained a letter from the NCES project officer, an informational brochure, any relevant endorsement letters from the National Catholic Educational Association (NCEA) or the National Association of Independent Schools (NAIS), and a publication entitled "Uses of Data for the Education Longitudinal Study of 2002 (ELS:2002). It also contained a state and/or district endorsement letter, if provided.

Several days after the package was sent, the school was contacted by telephone. After determining the appropriate person with whom to speak, the recruiter provided details about the study and answered any questions. If the school agreed to participate, a school coordinator was identified. This person served as a point of contact at the school and was responsible for handling the logistical arrangements. 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. It was determined whether the type of parental consent used by the school was active (written) consent or passive (implicit) consent. Schools were offered the opportunity to provide endorsement letters to be included with the consent letter to the parents. Among the participating schools, 114 (or about 15 percent of the sample) provided these letters.

The most common objection voiced during the recruitment process was concern about burden, loss of instructional time, and overtesting of students. These were the overwhelming reasons cited for refusals both at the district and school level.

In addressing the concerns, flexibility in scheduling was offered to the schools. Survey Days were conducted from mid-January through the beginning of June so that schools could choose a date when they were less busy. Some 61.6 percent (unweighted) of eligible schools participated and 38.4 percent refused. Of the school refusals, approximately 36 percent occurred at the district level and 64 percent at the school level. Eleven schools allowed administration of the student questionnaire but did not allow any testing. In 2 additional schools, the school allowed administration of the student questionnaire and a math test, but no reading test.

[47] An additional 14 districts were contacted. Ten districts reported information indicating that their selected school(s) were ineligible. Four districts were contacted conditionally, but their schools were not selected.
[48] There were 162 eligible sample schools not affiliated with districts/dioceses.

4.2.3 Pre-data Collection Activities

After obtaining school approval for the study, a study notebook was sent to the coordinator that detailed the tasks for which he/she was responsible and included instructions on preparing and sending a 10th-grade enrollment list. 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, ethnicity, and whether the student had an individualized education program (IEP). Some schools also agreed to provide Social Security numbers and/or school identification numbers to facilitate tracing in the longitudinal follow-up. After the enrollment list was received, students were sampled. The list of sampled students was sent to the coordinator and he/she was asked to provide address and telephone information for each student. The coordinator was also asked to provide the titles of the students' English and mathematics courses and the teachers of those subjects for each student. Approximately 2 months prior to the scheduled Survey Day, the coordinator was asked to send another enrollment list of their 10th-grade students. This information was used to identify students who had enrolled after the original list was provided.

Approximately 135 survey administrators were trained to conduct data collection in the schools. Prior to training, each survey administrator (SA) was mailed a copy of the survey administrator manual and a home study exercise. The SAs were instructed to read the manual prior to training and complete the home study exercise to be turned in on the first day of training. Training was held for 2 days each in Durham, North Carolina, and Los Angeles, California, in early January 2002. Staff from five field supervisor regions were trained at each session. With the exception of an introductory session that was held with the regional training group as a whole, the SAs were divided into training groups by field supervisor region. Each training room contained a lead trainer from the project staff and a field supervisor. The training agenda is shown below in figure 7.

Each SA received case assignment cards for each of his/her assigned schools. The case assignment cards contained information about the school, including the name and phone number of the school coordinator and the designated Survey Day and make-up day. After training, the SAs contacted each study coordinator prior to Survey Day to make logistical arrangements. These arrangements included verifying that the test supplies had arrived, that the coordinator had reserved a room for testing, and the coordinator had distributed staff questionnaires as well as reminder notices to sampled students. At the same time, the SA determined if the coordinator had received any parental refusals. If so, the SA began refusal conversion efforts (if the school was willing to provide a telephone number for the parent). In active (explicit) consent schools, the SA also determined from the coordinator which parents had not yet returned permission forms. If the school was willing to provide telephone numbers, the SA began calling the parents to prompt them to return the forms.

Figure 7. Survey administrator training agenda: 2002

Day 1		Day 2	
8:30 – 8:40	Introductions	8:30 – 9:15	Survey Day from start to finish
8:40 – 8:45	Confidentiality	9:15 – 10:30	Disposition of forms
8:45 – 9:15	Prior NCES studies/Overview of ELS:2002	10:30 – 10:45	BREAK
9:15 – 9:30	Prior contacts with schools	10:45 – 11:15	Contacting parents
9:30 – 10:30	Case assignment card (CAC), working with the school coordinator	11:15 – 11:30	Student and staff nonresponse follow-up
10:30 – 10:45	BREAK	11:30 – 12:00	Hiring and training survey administrator assistants (SAAs)
10:45 – 12:00	Working with the school coordinator (continued)	12:00 – 1:00	LUNCH
12:00 – 1:00	LUNCH	1:00 – 1:15	Dealing with disruptive students/ other problems at schools
1:00 – 2:00	Survey Day logistics	1:15 – 3:00	Headway procedures
2:00 – 2:15	Routing test exercise	3:00 – 3:15	BREAK
2:15 – 2:45	Student questionnaire	3:15 – 5:00	Distribution of assignments
2:45 – 3:15	Edit exercise		
3:15 – 3:30	BREAK		
3:30 – 4:00	Facilities checklist		
4:00 – 4:15	Staff questionnaires		
4:15 – 5:00	Make-up days		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Each SA recruited, hired, and trained a survey administrator assistant (SAA) to help in the school. In some cases, the SA was able to use the same SAA for all of the assigned schools. However, in a few cases, the schools were located far enough away from where the SA lived that conducting Survey Day involved an overnight stay. In that case, the SA hired an SAA who lived close to the school.

4.3 Data Collection Procedures, Student Questionnaires, and Tests

Prior to questionnaire and test administration, parents were sent consent letters to notify them about the study. As mentioned previously, during the recruitment process, the parental permission procedure was discussed with the schools. Schools were encouraged to allow passive (implied) consent unless the school expressed the need for active consent. A total of 95 schools (12.6 percent) required active parental consent.

For schools that required active parental consent, information packets were sent via Federal Express to all parents for whom street addresses were given. If only a post office box address was available, packets were sent via regular mail (since Federal Express only delivers to street addresses). For students without a provided mailing address, parent packets were sent to the school for the coordinator to distribute. Each packet contained a letter about the study, a consent form, a brochure about the study, and an envelope bearing the school coordinator's name (to whom parents could return the consent form). In some cases, the principal had drafted an endorsement letter that was also included. The packets were sent to coordinators 4 weeks prior to each school's scheduled Survey Day. Prior to Survey Day, the SAs checked with the coordinators to obtain the names of parents who had not yet sent back a consent form. If they

were given telephone numbers, the SAs telephoned the parents to prompt them to return the forms.

Very few parents returned forms expressing refusal to let the student take part. However, many parents did not return the form at all. As a result, only 1,335 of the 2,150 eligible 10th-grade students (62.1 percent) sampled at schools requiring active permission took part in the in-school portion of the study. An additional 141 students at those schools were interviewed by telephone for an overall student participation rate (unweighted) of 68.7 percent at active consent schools.

For schools that allowed passive parental consent, letters were sent via first-class mail to all parents for whom mailing addresses were available. For those without a provided mailing address, parental packets were sent to the school for the coordinator to distribute. The packets contained a letter about the study, a consent form, a brochure about the study, and an envelope bearing the school coordinator's name (for parents to use to return the consent form) parents could return the consent form. An endorsement letter from the school was included, if one was provided. Passive parent consent letters were sent 2 weeks prior to the scheduled Survey Day. SAs contacted the school coordinators prior to Survey Day to determine if any parents had sent back forms that refused consent. For those parents, the survey administrators attempted refusal conversion (if the school was willing to provide telephone numbers).

As occurred with the active consent schools, very few parents returned forms expressing refusal to let their students take part in the study. As a result, 13,494 of the 15,441 eligible 10th-grade students (87.4 percent) sampled from passive consent schools took part in the in-school portion of the study. An additional 392 students at those schools were interviewed by telephone for an overall student participation rate (unweighted) of 89.9 percent at passive consent schools.

Parental consent letters were available in English and Spanish. Both English and Spanish versions of the letter and study brochure were sent to parents of all students who had been identified as Hispanic by their schools. A version of the consent letter was translated into Mandarin, Vietnamese, Korean, and Tagalog. An English version of the letter and brochure was sent with the Asian language translations to parents of all students who had been identified as Asian by their schools.

Shortly before Survey Day, reminder postcards were mailed to the sampled students for whom addresses were available. School coordinators were also provided with reminder notices to distribute to the sampled students several days prior to Survey Day.

On the Survey Day at each school, the SA checked in with the school coordinator and collected any parental permission forms that had come in. In active consent schools, the SA checked the student roster to make sure that only students who had returned signed permission forms were allowed to participate. In both active and passive consent schools, the SA made sure that no one for whom the school had received a parental refusal was allowed to participate, unless the parent had rescinded that decision in writing. As students entered the testing room, they were checked off on the student roster. After the majority of the sampled students arrived, the school coordinator was asked to try to locate the students who were not present.

The SA and SAA administered the student questionnaire and tests via a group administration. First, students were given a timed routing test in math and reading. After completing the routing tests, the students completed the student questionnaire. While the students completed the questionnaire, the SA and SAA graded the routing tests. This was done by using an answer key that overlaid the test form. The SA used the scores from the routing test to determine the second-stage tests in math and reading (low, medium, high ability) to assign to each student based on ability level. After the questionnaires were collected, the SA gave the students a short break and served a light snack. After the break, the SA handed out second-stage cognitive tests. While the students completed the second-stage tests, the SA and SAA edited the student questionnaires for completeness. If a student neglected to answer a questionnaire item deemed to be critical (student questionnaire critical items are listed in appendix C), the SA/SAA asked the student to complete it after the end of the second-stage test.

If less than 100 percent of the eligible students participated on Survey Day, the SA attempted to confirm the make-up day that had been scheduled during the school recruitment process. Staff asked to return for a second make-up day if attendance was still below 100 percent at the first make-up day. Because of the reduced number of students participating, make-up days were staffed by one person (either the SA or the SAA) instead of two. In some cases, schools did not permit the SA to conduct a make-up day session. Of the 15,362 participants, 85.4 percent were surveyed in their school on survey day, another 11.1 percent were surveyed on make-up day, and 3.5 percent were surveyed outside school over the telephone.

Computer-assisted telephone interviews (CATI) were conducted for students who were unable to participate in the group-administered sessions. Student phone numbers were obtained from the school (when given), or via address tracing (if the school released an address only). Prior to conducting student telephone interviews, verbal parental consent was obtained.

When the response rate was likely to be depressed, students were offered gift certificates for participation. In schools that required active consent, a Survey Day drawing was held for two $20 gift certificates. In schools that only allowed survey administration during off-school hours (after school, weekends, or school holidays), each participating student was offered a $20 gift certificate. All participating students were also offered a $20 gift certificate when schools would not release student addresses and sent the parental consent materials out themselves.

4.4 Data Collection Procedures, Teacher Survey

The teacher questionnaire was designed to obtain teacher reports of information. As mentioned previously, at the time that the student sample was selected, the school coordinator was asked to provide the names of each sampled student's math and English teachers. The coordinator was asked for the name of the fall teacher if the student was enrolled in class during the fall. If the student was not enrolled in class in the fall, the coordinator was asked for the name of the spring teacher. Teacher data collection was conducted via a mailed questionnaire. Questionnaire packets were prepared for each teacher, and all of the packets were mailed to the school coordinator for distribution. Each packet contained a lead letter, a brochure explaining the study, the ELS:2002 Uses of Data booklet, a list of sampled students for that particular teacher, the teacher questionnaire, and a postage-paid return envelope. Teachers who were being asked to report on more than 16 students also received a supplementary teacher questionnaire.

Teachers were sent a reminder postcard that asked them to complete the questionnaire and return it. Prompting telephone calls were made to nonresponding teachers through September 2002.

If it was determined during prompting calls that a particular teacher had not taught the students identified as belonging to them, an attempt was made to identify the student's correct teacher and to send that teacher additional materials. During these calls, some teachers who had already returned questionnaires were identified. For those teachers, a supplemental questionnaire that contained only questions about individual students was mailed.

Incentives were offered to responding teachers, based upon the number of students that each teacher was asked to report on. Incentives offered were: $10 to teachers reporting on 1-5 students, $20 to those reporting on 6–10 students, $30 to those reporting on 11–15 students, and $40 to teachers reporting on 16 or more students.

By the end of the data collection period, at least one teacher report had been received for 92.4 percent of all of the participating students.

4.5 Data Collection Procedures, School Administrator Survey

At the time that the schools were recruited, the school coordinator was asked to designate an individual at the school to be responsible for completing the school administrator questionnaire. At the time that Survey Day materials were sent to the school, a packet for the person designated to receive the school administrator questionnaire was included. The packet contained a lead letter, a brochure explaining the study, the ELS:2002 Uses of Data booklet, the school administrator questionnaire, and a postage-paid return envelope. Because the bulk of the questions in the questionnaire were of a general nature about the school and its policies, any knowledgeable staff member was permitted to complete the majority of the questionnaire. It was required that the final section be filled out by the school's principal. Burden on the principal was reduced by the length of this section (it took about 5 minutes to complete) and the fact that someone else at the school could complete the rest of the questionnaire.

Prompting for school administrator questionnaires was done during contacts with the schools. A total of 663 questionnaires were received by mail (88.2 percent) and an additional 80 school administrators completed abbreviated questionnaires by telephone (10.6 percent) for a 98.8 percent (unweighted) administrator response rate. Completed school administrator questionnaires provide 99.0 percent (weighted) coverage of all responding students.

In an effort to determine the characteristics of schools that did not participate in ELS:2002, such schools (or their affiliated districts) were asked to complete a school characteristics questionnaire for nonresponding schools. This questionnaire gathered information about basic characteristics of the refusing schools, which were also asked for in the school administrator questionnaire for participating schools. Questionnaires were mailed to schools or districts and followed up by telephone as needed. Among the 469 nonresponding eligible sample schools, a total of 437 completed questionnaires (93.2 percent) were received.

4.6 Data Collection Procedures, Library Media Center Survey

At the time that the schools were recruited, the school coordinator was asked to designate an individual at the school to be responsible for completing the library media center questionnaire. This could be anyone on staff who was knowledgeable about the library media center. At the time that the Survey Day materials were sent to the school, a packet for the person designated to receive this questionnaire was included. The packet contained a lead letter, a brochure explaining the study, the ELS:2002 Uses of Data booklet, the library media center questionnaire, and a postage-paid return envelope.

Prompting for library media center questionnaires was done during contacts with the schools. A total of 718 questionnaires were received by mail (95.5 percent, unweighted). Completed library media center questionnaires represented 96.4 percent (weighted) coverage of all responding students.

4.7 Data Collection Procedures, Facilities Checklist

In addition to reports from students and staff about each school, there was also interest in obtaining an objective reporting about the physical plant. The facilities checklist was to be completed by the SA based on his/her observations in the building on the school's Survey Day. (Survey Days were normally held on Tuesdays, Wednesdays, or Thursdays.) The form was designed to be completed by the SA without assistance from school personnel. To achieve a measure of standardization in the observations, SAs were instructed to complete the form at the same time of day for each school. Most survey administrations took place in the morning; facilities checklists were completed immediately after the morning administration. However, in those cases in which there was an afternoon survey administration, the facilities checklist was completed prior to the survey administration. Procedures included reporting on conditions visible from the school's front hallway, student bathrooms, one classroom, the school's parking lot, and adjacent property. SAs were also asked to report on various security measures observed throughout the school building.

Survey administrators completed facilities checklists in all 752 schools.

4.8 Data Collection Procedures, Parent Survey

At the time that the ELS:2002 sample was selected from the school enrollment list, each school was asked to provide home addresses for the parents of each sampled student. In many cases, the schools provided addresses for all sampled students. In a few cases, schools provided addresses if they had a signed release on file for the student. In those cases, some but not all of the addresses were provided for sampled students. In other cases (specifically, in 14 schools, or about 2 percent of the school sample), the school would not provide any home addresses.

Parent questionnaires were mailed on the school's scheduled Survey Day to all parents for whom addresses had been provided. For parents with no address available, the parent questionnaire was not mailed until the student questionnaire was sent in and the locator information was recorded.

Parent questionnaire packets contained a lead letter and brochure explaining the study, the parent questionnaire, and a postage-paid return envelope. Packets were addressed "To the Parent/Guardian of [student's name]." Questionnaire instructions asked for the parent who was most knowledgeable about the child's education to complete the questionnaire. Questionnaires were available in English and Spanish.

One week after each parent questionnaire was mailed, a thank you/reminder postcard was sent. The postcard thanked the parents who had already completed and returned the questionnaire and asked those who had not to do so. Four weeks after the initial questionnaire mailing, the process of contacting nonresponding parents by phone and asking them to complete the survey by computer-assisted telephone interview (CATI) was begun.[49] For parents who expressed reluctance to participate, an abbreviated telephone interview to gather critical items was offered (if refusal conversion attempts proved to be unsuccessful).

About 1 month prior to the end of data collection, an abbreviated parent questionnaire was mailed to parents of participating students who had not yet responded. Parents were offered the option of completing the abbreviated questionnaire and returning it by mail or calling a toll-free number to complete the questionnaire by telephone interview. Of the 15,362 responding students, parent data (either by mailed questionnaire or by telephone interview) were received from 13,488 of their parents. This represents a weighted coverage rate of 87.4 percent.

4.9 Data Collection Results

Tables 43–45 summarize the data collection results for the ELS:2002 base year. Table 43 reviews the school sample selections and sample realization. The final sample size (752) was below the original target (800), for a 94 percent rate of sample realization. About 64.7 percent (631) of the initially fielded schools (976) cooperated; cooperation rates were lowest for the other private school sector.[50]

Table 44 displays weighted and unweighted completion rates based on the overall study/sample design, in which student questionnaire completers constitute the basic unit for the public-use files (students who, for reasons of English language limitations or disability, would have been unable to complete or validly complete the research instruments, were nevertheless included in the study; however, these students appear only on the restricted-use files). For purposes of this table, the completion rate was calculated as the ratio of the number of completed interviews divided by the number of eligible sample members. Note that the participating student sample defines the eligible parent and teacher samples. Teacher and parent reports appear on the public-use files only if they can be linked to a participating student.

[49] English-language parent interviews were conducted by CATI; however, Spanish-language parent telephone interviews were conducted with paper-and-pencil methods.

[50] As may be seen in table 41, sample realization was lowest for other private schools (with a target of 100, only 77 were recruited, compared to 95 recruited schools and a target of 100 in the Catholic school sector, and a target of 600 with 580 schools recruited for the public school sector). Cooperation was also lowest for the other private school sector, as may be seen in table 43. Weighted response rates were 69 percent for public, 74 percent for Catholic, but only 63 percent for other private schools. In contrast, in NELS:88, other private schools had the highest cooperation rate and public schools the lowest.

Table 45 shows weighted and unweighted participation rates for school recruitment, and response rates for the school components (administrator, library, facilities) at the school level. Overall 1,221 schools were selected and found to be eligible. Some 752 participated. Facilities checklists were collected at all 752 schools, library media center questionnaires at 718 of the schools, and school administrator questionnaires at 743 of the schools.

In considering participation rates, it is important to note that while school-level and individual-level response rates are often considered separately, effects of nonresponse in a two-stage sample are, for many purposes, multiplicative across the two stages. A true indication of the response rate for students can be computed by multiplying school participation rates by individual participation rates. For example, defining school participation in terms of the percentage of schools that held Survey Days, and multiplying that percentage by the overall student response rate, the overall response rates are:

- 59.2 percent (0.68*0.87) for students;
- 66.8 percent (0.68*0.99) for school administrators;
- 65.0 percent (0.68*0.96) for libraries; and
- 67.8 percent (0.68*1.00) for facilities checklist.

As a point of comparison, these multistage participation rates are similar to those of the 1980 HS&B base year survey and to those of NELS:88 base year.

Table 43. ELS:2002 base year school sample selections and realization: 2002

Stratum	Estimated size	Eligible original selections	Target *N*	Total *N* cooperating schools	Sample realization (percent of target achieved)	Cooperating original selections	Cooperating alternative selections
Total	24,397	976	800	752	94.00	631	121
Public schools	17,311	735	600	580	96.67	484	96
Catholic schools	1,098	117	100	95	95.00	83	12
Other private schools	5,988	124	100	77	77.00	64	13

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 44. ELS:2002 base year completion/coverage rates, by component at the student level: 2002

	Student questionnaire completion rates		Student test coverage rates		Parent questionnaire coverage rates		Teacher questionnaire coverage rates		School administrator's questionnaire coverage rates		Library media center coverage rates	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Total	87.28	87.33	95.08	94.67	87.45	87.80	91.64	91.66	98.98	99.04	96.37	95.84
Participated	15,362		14,543		13,488		14,081		15,215		14,723	
Selected	17,591		15,362		15,362		15,362		15,362		15,362	
School type												
Public	87.05	86.72	95.10	94.54	87.25	87.37	91.41	90.85	98.91	98.80	96.48	96.17
Catholic	90.26	90.87	96.73	95.99	91.99	90.94	96.92	96.93	100.00	100.00	94.35	95.21
Other private	89.86	87.91	92.50	93.94	87.42	87.17	91.30	91.38	99.54	99.79	95.69	93.87
Urbanicity												
Urban	85.01	85.42	94.99	94.25	85.14	86.92	88.11	88.86	98.42	98.79	94.01	93.49
Suburban	87.41	87.43	95.62	95.49	88.11	87.96	92.30	92.28	99.43	99.28	97.27	96.84
Rural	90.41	90.70	93.84	93.29	89.34	88.97	95.40	95.08	98.68	98.88	97.67	97.47
Region												
Midwest	87.68	88.66	95.75	94.53	85.79	87.11	95.14	95.31	99.94	99.92	98.83	96.83
Northeast	82.21	82.77	95.91	96.13	87.93	87.04	89.47	89.36	97.81	98.52	92.66	92.83
South	89.87	89.95	94.45	94.59	88.03	88.39	93.38	93.40	99.29	99.49	97.17	97.23
West	87.07	85.37	94.65	93.67	87.96	88.28	87.15	85.94	98.45	97.60	95.58	94.74
Race/Ethnicity[1]												
Asian/Pacific Islander	83.89	83.60	94.41	93.74	87.29	87.48	86.10	85.68	96.30	96.75	94.29	94.04
Black	87.02	86.91	95.62	95.29	83.12	83.90	89.66	90.29	97.81	98.37	93.25	93.56
Hispanic	85.74	86.16	95.92	95.77	86.86	87.12	86.28	88.63	98.29	98.32	94.20	94.49
Native American/Alaska Native	91.76	91.16	92.33	94.03	83.13	82.09	92.74	91.79	100.00	100.00	99.64	99.25
White	88.55	89.24	95.37	95.71	88.70	89.33	95.50	95.65	99.49	99.67	97.61	97.29
Other[2]	81.42	82.69	90.72	87.37	87.75	86.02	75.66	80.55	99.78	99.79	96.98	93.90
Highest parent education[3]												
Did not finish high school	†	†	96.24	95.06	82.74	82.60	88.23	88.08	98.52	97.85	93.85	92.48
High school graduate or GED[4]	†	†	95.23	94.48	83.31	83.19	91.81	90.84	99.02	99.05	96.92	96.09
Some college (< 4-year degree)	†	†	95.13	94.89	89.10	89.72	92.01	92.10	98.91	99.07	96.55	96.02
Bachelor's degree	†	†	95.60	95.41	88.43	88.75	91.77	91.99	98.88	99.00	96.12	95.78
Graduate/professional degree	†	†	93.61	93.44	89.79	89.87	91.84	92.53	99.42	99.44	96.58	96.42

† Not applicable.
[1] Race/ethnicity classification was based on school-provided sampling information.
[2] Other category includes multiracial and missing.
[3] Highest parent education was imputed if otherwise missing.
[4] GED = Graduate equivalency diploma.
NOTE: Facilities checklist coverage rates were 100 percent and do not appear in the table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table 45. ELS:2002 base year school-level participation rates and completion rates for school surveys at the school level: 2002

	School participation rates		School administrator questionnaire completion rates		Library media center completion rates		Facilities checklist completion rates	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Total	67.80	61.59	98.53	98.80	95.93	95.48	100.00	100.00
School type								
Public	69.09	62.63	99.62	98.62	97.43	95.86	100.00	100.00
Catholic	74.04	67.86	100.00	100.00	93.23	95.79	100.00	100.00
Other private	62.94	49.68	94.77	98.70	91.72	92.21	100.00	100.00
Urbanicity								
Urban	67.27	60.39	99.39	98.40	94.47	92.80	100.00	100.00
Suburban	59.81	59.28	99.86	99.45	98.23	96.95	100.00	100.00
Rural	79.32	71.21	96.62	97.87	94.37	96.45	100.00	100.00
Region								
Midwest	73.87	68.61	95.82	99.47	93.51	96.81	100.00	100.00
Northeast	60.37	52.14	99.24	97.76	94.07	91.04	100.00	100.00
South	70.33	72.87	99.75	99.29	98.64	97.16	100.00	100.00
West	63.06	48.84	99.64	97.97	96.11	94.59	100.00	100.00

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Chapter 5
Data Preparation and Processing

This chapter describes the automated systems used to control survey processes; procedures used to maintain receipt control; aspects of data preparation (such as coding); and the various procedures for data capture, cleaning, and editing. The chapter is organized into seven sections: (1) overview of systems design, development, and testing, (2) data receipt, (3) coding for hardcopy instruments, (4) data capture for optically scanned instruments, (5) data cleaning and editing, (6) data capture and editing for computer-assisted telephone interviewing (CATI), and (7) data processing and file preparation.

5.1 Overview of Systems Design, Development, and Testing

Most systems were designed during the field test with concern for the processes needed for the main study. The effort was to test systems in a smaller environment to reveal points in which improvements could be implemented on a larger scale. After the field test, improvements were implemented and checked in a test environment. The following systems were developed during the field test:

- a recruiting system;
- a Survey Control System (SCS);
- Survey Day materials generation programs;
- survey administrator telephone touch-tone data capture systems;
- a questionnaire receipt application;
- TELEform (application used for scanning questionnaires);
- a Structured Query Language (SQL) server database to store scanned data responses;
- a scanned image database;
- a parent computer-assisted telephone interview (CATI);
- parent CATI and scanned data concatenation programs;
- computer-assisted data entry (CADE) programs for the facilities checklist;
- data cleaning programs;
- a web-based Integrated Management System (IMS); and
- production reports.

A full development process, including design, programming, testing, and implementation was used in the creation of these systems. Specifications were developed in word processing documents and flowchart applications and progress was tracked using Microsoft Project and Microsoft Excel. Specifications for questionnaires were designed in word processing documents

and were updated to reflect what changed between the field test questionnaires and the full-scale questionnaires. CATI specifications were developed from the questionnaire specifications and included CATI application pseudo code for each CATI item.

Between the field test and full-scale studies, systems and procedures were evaluated and the following functionality was added to the full-scale operations:

- a Survey Day materials printing application (based on materials processed for Survey Day mailing);
- a mail return application;
- an incentive tracking application;
- a dynamic reporting utility that provided managers with current information from the SCS database;
- a student CATI component (linked to the parent CATI);
- a scanned image archive application that allowed images to be burned to compact disc (CD) archives;
- a scanned image archive server that allowed instant access to scanned questionnaires during the data cleaning and review process;
- a cleaning and editing application that allowed editors to review and correct questionnaire data as appropriate, working in conjunction with actual scanned images in cases in which inconsistent data occurred;
- a data review system that allowed reviewers to randomly review questionnaires with data in order to detect data deficiencies (e.g., scanning problems); and
- an occupation coding application.

5.2 Data Receipt

The data preparation facility received all materials returned to RTI after a school's survey was complete or individual respondents (e.g., school faculty, parents, etc.) sent in completed questionnaires. Procedures were established to systematically receive and record all required forms; this process included the scanning of bar-coded labels. Additional receipt events were added for the full-scale study to identify questionnaires that were not completed fully or accurately and to allow project staff to follow up promptly. Different versions of questionnaires (e.g., full, abbreviated, Spanish, etc.) were easily distinguishable within the receipt process and were automatically batched separately based on the questionnaire type. For example, Spanish questionnaires were translated to an English questionnaire in preparation for scanning.

After questionnaires were received and added to the receipt system, a batch number was assigned to the questionnaire. To assist the project team in cases that required referring to a questionnaire, the system was able to access dynamically the status of an individual questionnaire and provide the batch number that it belonged to. If the questionnaire had moved beyond the scanning stage, the scanned image could be accessed as well. Questionnaires were

occasionally identified for data removal (e.g., when parent consent was lacking). Rather than deal with the removal process manually, a spreadsheet was developed to document these cases and case removal was integrated into the data delivery process. This was a useful model because it did not disrupt the questionnaire processes and provided the ability to add cases back to final data files when appropriate (e.g., when parental permission was obtained).

5.3 Coding for Hardcopy Instruments

The following text items were obtained in the questionnaires:

- respondent occupation (from parent questionnaire);
- respondent partner occupation (from parent questionnaire);
- mother/female guardian occupation (from student questionnaire);
- father/male guardian occupation (from student questionnaire);
- expected occupation after high school (from student questionnaire); and
- expected occupation at age 30 (from student questionnaire).

The parent questionnaire allowed respondents to choose from a list of 16 occupation codes, relating to their occupation text. When occupation codes were not selected, the text was available for review and coding. The student questionnaire only collected occupation text, and did not provide the occupation categories.

Occupation text was loaded into a coding application (when occupation codes were lacking) in which a coding specialist could select the correct code from the 16 occupation categories. The resulting codes were merged back into the data files.

5.4 Data Capture for Optically Scanned Instruments

The following questionnaires were developed for optical scanning:

- a student questionnaire;
- an abbreviated student questionnaire;
- a first-stage routing test;
- second stage math and reading tests;
- a parent questionnaire;
- a school administrator questionnaire;
- library and media center questionnaires; and
- a facilities checklist.

Questionnaires were designed for TELEform scanning, and after questionnaires were received and batched they were ready for TELEform scanning. A TELEform questionnaire

contained text fields that could be recognized by scanning machines and interpreted forms text to data through optical character recognition. Verifiers reviewed data that was not interpreted accurately by the scanning machines or was not consistent with expected ranges. Once verification was complete, the data were converted to an American Standard Code for Information Interchange (ASCII) file and the questionnaire image was written to the server. This process provided immediate access to raw questionnaire data and a repository of images accessible by ELS:2002 staff.

TELEform development began with the field test TELEform document and specifications in Microsoft Word that indicated changes that were made between the field test and the full-scale study. Modifications were easily made and variable names were updated appropriately. Any new TELEform documents were first developed in Microsoft Word as a specification. As changes in the TELEform document were required, the corresponding Microsoft Word document was updated using the "Track Changes" tool. Reviewers would compare the specifications to the printed version of the TELEform document to ensure that all questionnaires were the latest version. When a TELEform document was confirmed as final, internal testing of the scanning and data-writing processes occurred. About 10 forms were printed and filled out for testing purposes. The test forms were scanned so that the resulting data could be compared to the original questionnaire; this comparison would detect problems with the printed questionnaire, the scanning program, or the Structured Query Language (SQL) server database.

Scanning procedures were evaluated after the field test in an effort to streamline the scanning process for the full-scale study. Different stages of the scanning process were timed, and averages across each stage (i.e., cutting, scanning, evaluation, verification, data/image commit) for each questionnaire were used to analyze system and staffing needs. The need for efficient archiving procedures arose from the large amount of space taken by scanned images on the server and the need for access to the image for review. An application was developed to control the archiving process across the tens of thousands of scanned images. Archive procedures were modified from those used during the field test and an SQL database was created to track what had been archived (and to which CD volume) for easy image retrieval.

Questionnaire data were committed to ASCII data files and loaded with a scheduled process into a SQL server database each night. Raw SQL server data were compared to the original questionnaires to ensure that scanning procedures were accurately storing data to the SQL server. The SCS tracked each form that was scanned by indicating a scanned event whenever the SQL Server database was updated for a questionnaire. If for some reason a record was not transmitted successfully before or during the commit (i.e., nightly loading process) to the SQL server, a scanned event would be lacking for the questionnaire and could be easily identified later for rescanning. This approach ultimately ensured that all questionnaires received would eventually have a corresponding data record and could not be dropped without detection.

5.5 Data Cleaning and 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 during frequency reviews, values out of expected ranges, and values not adhering to a particular skip pattern. The

review application provided the case/item level information, reasons for 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 a 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 documentation on recommended edits. Raw data were never directly updated, so changes were always stored cumulatively and applied each time a cleaned data set was produced. This provided the ability to provide documentation on 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 to either correct previously answered (or unanswered) questions to match the dependent items or blank out subsequent items to stay consistent with previously answered items.

5.6 Data Capture and Editing for CATI

In an effort to boost response rates, the following CATI instruments were developed: student (developed from the TELEform abbreviated version) and parent (developed from the TELEform full-length version, plus a module to administer an abbreviated version).

CATI logic was designed such that the TELEform and CATI records could be concatenated into one data file. It is possible that a respondent was included in both the TELEform and CATI sample; however, rules were implemented to identify these cases during file concatenation. The TELEform source took precedence over the CATI source unless the TELEform source was incomplete.

CATI instruments were developed with logic based on the skip patterns in the questionnaires. Questions were automatically skipped during administration. The questionnaire development program (CASES) stored data for each item answered, but respondents were allowed to go back to previously answered items. In rare cases, a previously answered item could be changed in such a way that the questionnaire logic was inconsistent with data already answered from a different logical path. Editing programs were built to review and ultimately blank out items that would not have been answered otherwise.

5.7 Data Processing and File Preparation

All TELEform questionnaire scans were stored in a SQL server database. CATI applications were used to supplement questionnaires where Paper and Pencil Interviewing (PAPI) was not always desirable. CATI data were exported nightly to ASCII files. Cleaning

programs were designed to concatenate CATI and TELEform SQL Server data into SAS data sets, adjusting and cleaning variables when formats were not consistent. Special attention focused on this concatenation to verify that results stayed consistent and to rule out possible format problems. In some cases, data were collected from both modes of administration for a respondent. Procedures were developed to remove the duplication within the raw data sets by selecting the latest scanned case; however, this rule was overridden with alternative records when record-by-record comparison determined otherwise.

Once questionnaire data were concatenated and cleaned across modes and versions, the following cleaning and editing steps were implemented:

- anomalous data cleaning based on review of data with original questionnaire image (e.g., scanning errors);
- rule-based cleaning (changes that were made based on patterns in data, rather than review of images);
- hard-coded edits based on changes recommended by a reviewer if respondents misunderstood the questionnaire (e.g., respondent was instructed to enter a percentage; however, there was strong evidence that the respondent entered a count rather than the percentage); and
- edits based on logical patterns in questionnaire (e.g., skip pattern relationships between gate and dependent questions).

All respondent records in the final data set were verified with the SCS to spot inconsistencies. For example, it was possible that data were collected for a respondent who later was set to an ineligible status. It would not be appropriate to include that data, and the SCS served as a safeguard to ensure data integrity. Furthermore, the data files served as a check against the SCS to ensure that all respondent information was included in production reports.

Item documentation procedures were developed to capture variable and value labels for each item. Item wording for each question 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.

Chapter 6
Data File Contents

This chapter provides a concise account of the Education Longitudinal Study of 2002 (ELS:2002) base year data file contents. It addresses the following six topics: (1) structure of the Electronic Codebook (ECB) system, (2) analysis populations, (3) weights and flags, (4) composite and classification variables, (5) variable naming conventions, and (6) the hardcopy student component codebook.

6.1 Data Structure

ELS:2002 base year data have been made available in public- and (for licensed users) restricted-use versions[51] in an Electronic Codebook (ECB) format on CD-ROM. The ECB is designed to be run in a Windows environment. The ECB is available (at no cost) from the National Center for Education Statistics (NCES). Appendix A supplies a brief introduction to the ECB, including its installation.

The ECB system serves as an electronic version of a fully documented survey codebook. It allows the data user to browse through all ELS:2002 variables contained on the data files, to search variable and value names for key words related to particular research questions, to review the wording of these items along with notes and other pertinent information related to them, to examine the definitions and programs used to develop composite and classification variables, and to "output" the data for statistical analysis. The ECB also provides an electronic display of the distribution of counts and percentages for each variable in the data set. Analysts can use the ECB to select or "tag" variables of interest, print hardcopy codebooks that display the distributions of the tagged variables, and generate SAS and SPSS program code (including variable and value labels) that can be utilized with the analyst's own statistical software.

The ECB comprises two megafiles: first, a megafile at the student level, which encompasses student, parent, and teacher data; and second, a megafile at the school level, which encompasses data from the facilities checklist, the school administrator questionnaire, and the library media center questionnaire. Weights, participation flags and status indicators, and composite and classification variables come first on the file, followed by the questionnaire variables.

6.2 Base Year Analysis Populations

The base year data can only be used cross-sectionally at this time, as a description of America's high school sophomores as of the spring term of the 2001–2002 school year. However, its cross-sectional use includes cross-cohort (intercohort) comparisons with two earlier national samples of sophomores: High School and Beyond Longitudinal Study (HS&B) sophomores in 1980, and National Education Longitudinal Study of 1988 (NELS:88) sophomores in 1990. Appendix H includes a discussion of special issues in comparing the

[51] A license is required to access the restricted-use ECB.

cohorts and a crosswalk of common items among the three studies. Also, equated test scores have been generated so that achievement in reading and mathematics in NELS:88, and in mathematics in HS&B, can be compared across the three studies. An equated score has also been provided, putting reading scores (and in the future, math scores) of the ELS:2002 sophomore cohort and the 15-year-old cohort of the Program for International Student Assessment (PISA) on the same scale.

6.3 Weights and Flags

The public-use files contain one weight for use with student data (BYSTUWT) and one weight for use with school-level data (BYSCHWT). A further student weight (BYEXPWT), that encompasses all students in the study including those who were exempted from taking the survey instruments because of limited English proficiency or a severe disability, appears only on the restricted file, as well as the design (raw or base) weight from which the final weight was derived.

Participation flags (which are always dichotomous) and status variables (which have more than two values), as well as weights, may be used for subsetting—in other words, they can be used to select the subset of respondents that the analyst intends to examine. For example, if one wishes to select only those students for whom there are assessment data, the status variable BYTXSTAT would be invoked (a "0" means no assessments were completed; a "1" means a reading test only was completed; a "2" indicates a mathematics test only was completed; and a "3" indicates both tests [assessments in reading and in mathematics] were completed). If one wishes not to use the imputed test scores, then the imputation flag must be invoked, for example, BYMATHIM ("1" means a missing mathematics score was imputed, and "2" means that it was not).

6.4 Composite and Classification Variables

Composite variables—also called constructed, derived, or created variables—are usually generated using responses from two or more questionnaire items or from recoding of a variable (typically for disclosure avoidance reasons). Some are copied from another source (e.g., a variable supplied in sampling, or a variable imported from an external database). Examples of composite variables include school variables (school sector, urbanicity, region of the country), assessment scores (achievement quartile in reading or in math), psychological scales (mathematics self-efficacy), and demographic variables (sex, race, Hispanic ethnicity, and month and year of birth).

Most of the composite variables can be used as classification variables or independent variables in data analysis. For purposes of better estimation in cross-sectional analysis, many of the composites have undergone imputation procedures for missing data (all imputed versions of variables have been flagged).

6.5 Naming Conventions

Data users should find naming conventions for variables, flags, and weights intuitive and quite similar to those employed in NELS:88. Most variables begin with an indicator of the wave (in this case, the base year, BY). Weights follow the same wave-naming convention and also contain the suffix "WT" (e.g., BYSTUWT, is the name for the final student weight for base year questionnaire completion, BYSCHWT the name for the final school weight). First follow-up variables will begin with the prefix "F1," second follow-up with "F2," and so on. A few composite variables will be updated round by round, as new respondents (freshened students or prior round nonrespondents) enter the responding sample. These cross-wave composites (e.g., SEX, RACE) have no prefix indicative of wave, because they are round independent.

Variable names also distinguish (in their third character or third and fourth characters) between components. "BYS," for example, indicates a base year student questionnaire variable, while "BYP" stands for base year parent. Likewise "A" is used for the principal (school administrator) questionnaire, "TM" for reports from the mathematics teacher, "TE" for the English teacher, "L" for the library media center instrument, and "F" for the facilities checklist. Variables that reflect specific items in the questionnaire carry the question number in the variable name, immediately after the component indicator. Hence, BYS26 would be item 26 from the base year student questionnaire, and BYP41 would be item 41 in the parent instrument.

The round-specific constructed variables are typically not anchored in a single questionnaire item and may sometimes reflect nonquestionnaire sources of information, such as the assessments. Test scores carry the prefix BYTX. BYTXMQU, for example, indicates the quartile score for the base year mathematics test. Flags are indicated by the suffix "FLG" or "FG." Variable names also distinguish between the public (P) and restricted (R) use forms, where variables differ between them.

6.6 Guide to the Hardcopy Codebooks

Although for most purposes the flexibility of the electronic codebook will best meet users' needs, in some situations it may be helpful to also have access to a specialized hardcopy codebook of the student data. The hardcopy codebooks appear as PDF files only for the web-published version of this manual (see *http://nces.ed.gov/surveys/els2002*), and corresponds to appendix G of this document. It supplies a comprehensive description of the student data file. For each variable on the student component data file, the codebook provides a summary of the related information, including the question number and wording, variable name, and the responses to the item along with their unweighted frequency and percent and weighted percent. It also provides missing data frequencies sorted by the following reserve codes:

- "Don't know"[52]
- "Refused"
- "Multiple response"
- "Not reached"

[52] For the sake of convenience, "Don't knows" receive a common reserve code, but in hardcopy codebooks and other contexts as well, a distinction is made between "Don't know" arising from a response volunteered in a CATI interview (a true reserve code), and "Don't know" arising from a legitimate response option in a questionnaire (which need not be looked at as a true reserve code).

- "Legitimate skip/NA"
- "Nonrespondent"
- "Out of range"
- "Data suppressed"
- "Missing"

Information on obtaining the ELS:2002 ECB (and other NCES electronic codebooks) can be found by reviewing the data products for the study at *http//nces.ed.gov/pubsearch*. Information on applying for a restricted-use license also appears on the NCES web site: *http://nces.ed.gov/pubsearch/licenses.asp*

References

Barton, P., and Coley, R.J. (1990). *The Education Reform Decade*. Princeton, NJ: Educational Testing Service.

Burns, L.J., Heuer, R., Ingels, S.J., Pollack, J.M., Pratt, D.J., Rock, D., Rogers, J., Scott, L.A., Siegel, P., and Stutts, E. (2003). *ELS:2002 Base Year Field Test Report* (NCES Working Paper 2003–03). U.S. Department of Education, National Center for Education Statistics.

Campbell, J.R., Hombo, C.M., and Mazzeo, J. (2000). *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance* (NCES 2000–469). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Carlson, B.L., Johnson, A.E., and Cohen, S.B. (1993). An Evaluation of the Use of Personal Computers for Variance Estimation with Complex Survey Data. *Journal of Official Statistics*, *9*(4): 795–814.

Chromy, J.R. (1981). Variance Estimators for a Sequential Sample Selection Procedure. In D. Krewski, R. Platek, and J.N.K. Rao (Eds.), *Current Topics in Survey Sampling* (pp. 329–347). New York: Academic Press.

Cohen, S.B. (1997). An Evaluation of Alternative PC-Based Software Packages Developed for the Analysis of Complex Survey Data. *The American Statistician*, *57*(13): 285–292.

Cox, B.G. (1980). The Weighted Sequential Hot Deck Imputation Procedure. *Proceedings of the Section on Survey Research Methods* (pp. 721–726). The American Statistical Association.

Curtin, T. R., Ingels, S.J., Wu, S., and Heuer, R. (2002). *NELS:88 Base Year to Fourth Follow-Up Data File User's Manual* (NCES 2002–323). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Deville, J.C., and Särndal, C-E. (1992). Calibration Estimating in Survey Sampling. *Journal of the American Statistical Association*, *87*: 376–382.

Folsom, R.E., and A.C. Singh. (2000). The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification. *Proceedings of the Section on Survey Research Methods* (pp. 598–603). The American Statistical Association.

Folsom, R.E., Potter, F.J., and Williams, S.R. (1987). Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains. *Proceedings of the Section on Survey Research Methods* (pp. 792–796). The American Statistical Association.

Frankel, M.R, Kohnke, L., Buonanno, D., and Tourangeau, R. (1981). *High School and Beyond Sample Design Report*. Chicago: National Opinion Research Center.

Hambleton, R.K. (1989). Principles and Selected Applications of Item Response Theory. In R.L. Linn (Ed.), *Educational Measurement*, 3rd ed. (pp. 147–200). New York: MacMillan.

Hambleton, R.K., Swaminathan, H., and Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Hurst, D., Tan, A., Meek, A., and Sellers, J. (2003). *Overview and Inventory of State Education Reforms: 1990 to 2000* (NCES 2003–020). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J. (1996). *Sample Exclusion in NELS:88—Characteristics of Base Year Ineligible Students: Changes in Eligibility Status After Four Years* (NCES 96–723). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Burns, L.J., Chen, X., Cataldi, E.F., and Charleston, S. (2004). *A Profile of the American High School Sophomore in 2002* (NCES 2004–396). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Curtin, T.R., Kaufman, P., Alt, M.N., and Chen, X. (2002). *Coming of Age in the 1990s: The Eighth-Grade Class of 1988 12 Years Later* (NCES 2002–321). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M., and Rasinski, K. (1994). *NELS:88 First Follow-Up Final Technical Report* (NCES 94–632). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Scott, L.A., and Taylor, J.R. (1998). *NELS:88 Base Year Through Second Follow-Up Final Methodology Report* (NCES Working Paper 9806). U.S. Department of Education, National Center for Education Statistics.

Jones, C., Clarke, M., Mooney, G., McWilliams, H., Crawford, I., Stephenson, B., and Tourangeau, R. (1983). *High School and Beyond 1980 Sophomore Cohort First Follow-Up (1982) Data File User's Manual* (NCES 83–214). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Kish, L., and Frankel, M.R. (1974). Inference from Complex Samples. *Journal of the Royal Statistical Society*, Series B (Methodological), *36*: 2–37.

Lemke, M., Calsyn, C., Lippman, L., Jocelyn, L., Kastberg, D., Liu, Y., Williams, T., Kruger, T., and Bairu, G. (2001). *Outcomes of Learning: Results From the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy* (NCES 2002–115). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

McLaughlin, D., and Cohen, J. (1997). *NELS:88 Survey Item Evaluation Report* (NCES 97–052). U.S. Department of Education, National Center for Education Statistics, Washington, DC: U.S. Government Printing Office.

Rasinski, K., Ingels, S.J., Rock, D.A., and Pollack, J.M. (1993). *America's High School Sophomores: A Ten Year Comparison, 1980-1990* (NCES 93–087). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Riccobono, J.A., Henderson, L.B., Burkheimer, G.J., Place, C., and Levinsohn, J.R. (1981). *National Longitudinal Study: Base Year (1972) through Fourth Follow-Up (1979) Data File User's Manual*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Rock, D.A., and Pollack, J.M. (1995a). *Psychometric Report for the NELS:88 Base Year Through Second Follow-Up* (NCES 95–382). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Rock, D.A., and Pollack, J.M. (1995b). Mathematics Course-Taking and Gains in Mathematics Achievement (NCES 95–714). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Scott, L.A., Rock, D.A., Pollack, J.M., and Ingels, S.J. (1995). *Two Years Later: Cognitive Gains and School Transitions for NELS:88 Eighth Graders* (NCES 95–436). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Spencer, B.D. (1991). Eligibility/Exclusion Issues in the 1990 Trial State Assessment. In R. Glaser, R. Linn, and G. Bohrnstedt (Eds.), *Assessing Student Achievement in the States: Background Studies* (pp. 1–18). Studies for the Evaluation of the NAEP Trial State Assessment. Stanford, CA: National Academy of Education.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K., and Tourangeau, R. (1990). *NELS:88 Base Year Sample Design Report* (NCES 90–463). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). *The Psychology of Survey Response*. New York: Cambridge University Press.

Tourangeau, R., Sebring, P., Campbell, B., Glusberg, M., Spencer, B.D., and Singleton, M. (1987). *The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up (1986) Data File User's Manual* (NCES 87–406c). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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

Woodruff, R.S. (1971). A Simple Method for Approximating the Variance of a Complicated Estimate. *Journal of the American Statistical Association*, *66*: 411–414.

Zahs, D., Pedlow, S., Morrissey, M., Marnell, P., and Nichols, B. (1995). *High School and Beyond Fourth Follow-Up Methodology Report* (NCES 95–426). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Appendix A
Introduction to the Electronic Codebook

This page is intentionally blank.

Appendix A
Introduction to the Electronic Codebook

This appendix supplies a brief introduction to the ELS:2002 data in Electronic Codebook (ECB) format. Special attention is given to general instructions on how to get started using the ELS:2002 data and an orientation to ECB and variance estimation software that can be used to manipulate the data.

A.1 Obtaining the ELS:2002 ECB

The ELS:2002 base year ECB on CD-ROM carries the NCES publication number NCES 2004–404. This data product contains

- ELS:2002 data from the base year;
- ECB software; and
- documentation.

A single copy of an ELS:2002 public-use CD-ROM may be obtained without cost from the Education Publications Center (ED Pubs), until supplies are exhausted. This group can be contacted by telephone at 1–877–4ED–PUBS or by writing

ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398.

Requests can also be made electronically to *http://www.edpubs.org/* or to *customerservice@edpubs.org*. Requesters will need the title of the data product and the NCES number (NCES 2004–404 for the ELS:2002 ECB).

A restricted-use version of the ECB is available to institutionally based users in the United States whose research requires this additional level of information. Contact NCES at *http://nces.ed.gov/pubsearch/licenses.asp.*
.

A.2 Features and Content of the ELS:2002 ECB

ECBs allow the user to

- search a list of variables based on keywords or labels;
- tag (i.e., select) variables for analysis;
- generate SAS and SPSS syntax for system files;
- produce printed codebooks of selected variables;
- import tag files; and
- access database files for extraction.

The overall organization of data reflects two integrated and comprehensive data files, or megafiles. One megafile is at the student level, the other at the school level. School-level variables include information collected in the base year school administrator questionnaire, library media center questionnaire, and facilities checklist. At the student level, data from the student questionnaire, the student assessments in reading and mathematics, the teacher questionnaire, and the parent questionnaire are represented. Weights, participation flags and status indicators, and composite variables (also called constructed variables, derived variables, or created variables) are located at the beginning of the file, followed by the questionnaire variables.

Some important variable naming conventions (normally embedded in the first 3 to 4 characters of each variable name) may be noted. Normally the first 3 to 4 characters of each variable name identify the instrument from which the variable is taken. BYS stands for base year student; BYS21 stands for question 21 in the student questionnaire. BYP stands for base year parent, BYA for the base year administrator questionnaire, and so on. A label with the terminal characters "WT" is indicative of a weight (e.g., BYSTUWT is the final or nonresponse-adjusted student weight for the base year). Test variables contain the characters "TX," while flags are indicated by FLG or FG and status variables by ST (e.g., BYTXSTAT refers to test completion status in the base year). The content of the student and school megafiles is described more specifically in the sections below.

A.2.1 Student Megafile

The student-level file contains variables from the student, parent, and teacher questionnaires, as well as scores for the assessments in reading and mathematics.

The main contents of the student file, in order of appearance, and associated naming conventions, are as follows:

- *IDs and Weights*. Student and school IDs and weights (BYSCHWT, BYSTUWT) are at the beginning of the data file.
- *Student-level Composites*. Student-level composites are typically derived from student or parent sources. Included with student-level composites are BYTX* variables for data associated with the reading and math assessments.
- *School-level Composites*. School-level composites have been replicated at the student level for analytical convenience.
- *Data from Outside Sources*. The restricted-use ECB, but not the public-use ECB, includes access to CCD/PSS data, replicated at the student level, as well as confidential geocode data and linkages to external sources.
- *Imputation Flags (e.g., –IM, as in BYMATHIM)*. These flags indicate whether missing values for a variable or composite were imputed.
- *Participation Flags (e.g., –FLG, FG, or F, as in BYTEQFLG)*. These indicators are dichotomous. They indicate whether or not some feature of the data is available for a respondent (e.g., Spanish-language parent questionnaire, teacher ratings, etc.).

- *Status Flags (e.g., –STAT, as in BYTXSTAT).* These indicators have more than two values, but are otherwise similar to participation flags; they indicate the participation status of sample members and availability of contextual data for them.
- *Student Questionnaire Data (BYS*).* These data come from scanned forms filled out by the student or from the computer-assisted telephone interview (CATI).
- *Parent Questionnaire Data (BYP*).* These data come from scanned forms filled out by the parent or from the CATI interview.
- *English Teacher Data (BYTE*).* These data come from scanned teacher questionnaires filled out by the student sample member's English teacher. English teacher data have been linked to the appropriate student(s).
- *Math Teacher Data (BYTM*).* These data come from the scanned teacher questionnaire and have been linked to the appropriate student(s).

A.2.2 School Megafile

The school-level file contains all questionnaires administered at the school level. This includes the school administrator questionnaire, the library media center questionnaire, and the facilities checklist.

Variable prefixes on the school file identify the contents:

- *IDs and Weights*. Student and school identifications (IDs) and the school weight (BYSCHWT) are at the beginning of the data file
- *School-level Composites*. School-level composites are produced from questionnaire data allowing an analyst access to data in an easier format.
- *Data from Outside Sources*. Licensed users of the restricted-use file will have access to CCD/PSS data via NCES identification number (NCESID), geocodes, and other information for linking to external sources.
- *School Administrator Data (BYA*).* These data come from scanned forms filled out by the school principal and other administrative staff.
- *Library Section Data (BYL*).* These data come from scanned forms filled out by the librarian or library media center specialist.
- *School Facilities Data (BYF*).* These data come from scanned forms filled out by the survey administrator during the student surveys at the school.

The school ID is constructed such that student file records can merge with the school data.

A.3 Installing the ECB

A.3.1 Hardware/Software Requirements

The ECB program is designed to run on a PC with Windows 95 or higher versions.

A.3.2 Installation Procedures

To install the ECB, complete the following steps:

1. Close all applications on your computer.
2. Place the CD–ROM into the CD-ROM drive.
3. From Windows, click on "START" and then "RUN."
4. Browse through the CD-ROM drive for the "ecbw" folder and open "SETUP.exe" file.
5. Setup will guide you through the installation of the ECB.
6. Click on ECB icon to run.

A.4 Using the ECB

A.4.1 Understanding the File Structure and Capacity

The ECB is ready to use once it is installed. To understand quickly the structure of the file and the power provided by the ECB to produce data files requires an understanding of the "hot" keys and some practice:

1. On the toolbar found at the top of the ECB screen, click on each "hot" key.
2. Consult the "Electronic Codebook Help Guide" available on the CD-ROM (file named HELP.pdf) for a specific overview of the ECB functions.

A.4.2 Examining the Frequencies Available for Each Variable on the ECB

By examining these data descriptions, the ELS:2002 user will begin to appreciate the complexity of collecting data from respondents (legitimate values, legitimate skips, refusals, etc.). It is important to realize that some respondents

- did not respond to the entire instrument;
- skipped individual items;
- refused to complete selected items;
- did not reach the end of the questionnaire;
- completed abbreviated versions of the instrument;

- made illegal skips; and/or
- responded outside predefined valid ranges.

The following reserve code conventions are used in the ELS:2002 data files:

- –1 = *"Don't know."* There are some instances where respondents are allowed to answer "Don't know" for questions in the hardcopy questionnaires, and this reserve code will apply. The parent CATI interview by default allows "Don't know" for most questions that a respondent does not know so that the subsequent question can be administered.

- –2 = "*Refused."* Respondents are free to refuse to answer any question to which they do not wish to respond. In the hardcopy questionnaire, such refusals are explicitly captured only for critical items (items that, because of their importance, are subject to onsite edit and retrieval). CATI interviews, by default, allow refusals to be recorded on a question-by-question basis.

- –3 = *"Legitimate Skip/NA."* Questions that are not answered because prior answers route the respondent elsewhere will be filled with "Legitimate Skip/NA." This value applies to variables from all data collection modes.

- –4 = *"Nonrespondent."* "Nonrespondent" variables from questionnaires that have no respondent are filled with the "Nonrespondent" reserve code. This applies to both the student file and the school file because each file is composed of multiple interviews. For example, the school file may contain school administrator questionnaire data and facilities data, but the school's librarian may not have responded to the library media questionnaire; hence all library media variables appear with the "Nonrespondent" reserve code.

- –5 = *"Out of Range."* Values reported by the respondent that are out of range. Certain responses were set to this value if they were beyond the reasonable limits for the given item. For example, a teacher may have indicated teaching at a particular school for a longer period of time than he/she taught overall.

- –6 = *"Multiple Response."* Non-CATI applications do not have the ability to prevent respondents from answering multiple responses to a question that requires one answer. The scanning process for hardcopy questionnaires routed these instances to a verifier to determine whether the respondent "intended" to choose one answer (e.g., eraser marks interpreted by the optical scanning equipment as a second answer). In the case that the verifier cannot determine a single unique answer, the item was assigned the reserve code for "Multiple Response."

- –7 = *"Not Reached."* Questions that are not answered because the respondent does not wish to continue the interview or, in timed sessions, because they have run out of time, are filled with a "Partial/Not Reached" reserve code. This

code was also used for the parent CATI interviews that encountered break offs during the interview (and the respondent could not be reached for completion of the interview). This reserve code is also used in the instance of use of an abbreviated version of the questionnaire, in which particular items were not included.

- –9 = *"Missing."* Questions that are not answered within the scanned hardcopy questionnaires. These questions are typically missed accidentally (e.g., respondent did not understand the routing pattern) and are not an indication of the respondent filling out only part of the questionnaire. This reserve code can also apply to CATI data where, for reasons associated with different versions, an item was not administered.

A.4.3 Creating a Taglist, Extracting Data, and Generating Program Code

The following procedures can be used to tag variables, extract data, and generate program codes on the ECB:

1. Tag variables of interest by clicking on the "tag box" next to each variable.
2. Choose the appropriate weights and flags for the population of interest. In each megafile, flags can be selected to identify a particular part of the population. For example, flags are available to identify whether a student questionnaire completer also completed a test. Weights are variables placed on the dataset to compensate for the unequal probabilities of selection and to adjust for nonresponse. When used with flags, weights allow the analyst to make generalizations about the national populations represented by the various ELS:2002 samples (e.g., schools versus students within schools). When weights are not used and/or when a flag is used inappropriately, the estimates generated will not be representative of the population.
3. After tagging the variables of interest, go to "File" and then "Output."
4. Select the program (e.g., SPSS to generate SPSS program code).
5. Specify directory and name of program code file.
6. Select appropriate button in "Confirmation" box.
7. To view the program code, select "File" and then "View Output."
8. Open the program code in the appropriate software (e.g., SPSS) to generate a working system file and run analyses. It may be necessary to modify the program slightly (check for "execute" statements, period locations, and file names). The code should identify the ASCII data file location, which will be the CD-ROM. Users should be aware of a possible SPSS syntax error associated with continuous variables: the "VALUE LABELS" statement is missing when the first tagged item for a data file is continuous and has no reserve codes.

A.4.4 Variance Estimation

Because the ELS:2002 sample design involved stratification, disproportionate sampling of certain strata (e.g., oversampling of Asians and of private schools), and clustered (e.g., students within a school) probability sampling, the resulting statistics are more variable than they would have been had they been based on data collected from a simple random sample of the same size. A number of statistical packages (e.g., SUDAAN, WESVAR, STATA, and AM) take account of complex sampling designs in the calculation of standard errors. (For an assessment of strengths and limitations of SUDAAN, Stata, and WesVar, see Broene and Rust 2000.) AM variance estimation software can be downloaded for free from the following website: *http://am.air.org/*. While users are strongly urged to employ variance estimation software, an alternative that supports the generation of approximate standard errors is use of the design effect (for details, see chapter 3 of the ELS:2002 *Data File User's Manual*).

A.5 Additional Sources of Information (NCES Reports, Bibliographic Resources)

Although only one report using ELS:2002 data has been produced to date (*A Profile of the American High School Sophomore in 2002*, NCES 2003–396), many more are planned. In addition, many of the National Education Longitudinal Study of 1988 (NELS:88) reports may be of interest, both for what they suggest about possible cross-cohort analyses, and for issues that can be examined cross-sectionally in ELS:2002 and NELS:88. ELS:2002 reports can be found in electronic format on the NCES website under *http://nces.ed.gov/pubsearch/getpubcats.asp?sid=023*. From the NCES website, documents can be searched and downloaded.

To aid researchers in locating reports that use ELS:2002 data, NCES has contracted with RTI International to produce a comprehensive annotated bibliography of publications and reports (including doctoral dissertations) that draw on ELS:2002 data. While the ELS:2002 bibliography cites only a handful of sources at this time, the number will rapidly grow. In addition, a bibliography is actively maintained for NELS:88. This bibliography can be found at: *http://nces.ed.gov/surveys/nels88/*. The NELS:88 bibliography will be valuable for those who plan cross-cohort analyses using ELS:2002 and NELS:88 sophomore data, and may serve as well to suggest particular cross-sectional analyses that have proved fruitful in the past and may be worth pursing with the ELS:2002 data set.

A.6 Reference

Broene, P., and Rust, K. (2000). *Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets* (NCES 2000–03). U.S. Department of Education. Washington, DC: National Center for Education Statistics Working Paper.

This page is intentionally blank.

Appendix B
Base Year Questionnaires

This page is intentionally blank.

Appendix B
Base Year Questionnaires

Web-published PDF files of the base year questionnaires are available at: http://www.nces.ed.gov/surveys/els2002/index.asp

This page is intentionally blank.

Appendix C
Student Questionnaire Critical Items

This page is intentionally blank.

Appendix C
Student Questionnaire Critical Items

Critical items are data elements deemed to be of special importance (for future locating of the respondent, for research, or as a data quality check on whether skip patterns are being followed correctly) and therefore are subject to edit and retrieval in the course of the in-school survey session.

Table C–1. ELS:2002 student questionnaire critical items

Variable	Variable description
BYS01	Name, address, phone number♦
BYS02	Mother's name♦
BYS03	Is her phone number the same as respondent's?♦
BYS04	Mother's address and home telephone number♦
BYS05	Her work phone number♦
BYS06	Father's name♦
BYS07	Is his address and telephone number same as respondent's?♦
BYS08	Father's address and home telephone number♦
BYS09	Father's work telephone number♦
BYS10	Name address and telephone number of relative or close friend♦
BYS13	Date of birth
BYS14	Sex of student
BYS15	Hispanic ethnicity
BYS16	Hispanic subgroup
BYS17	Race
BYS18	Asian subgroup
BYS19	Social Security number♦
BYS56	How far in school expect to get
BYS57	Plans for continuing education after high school
BYS67	Is English the student's native language
BYS68	Native language if not English
BYS81	Mother's occupation
BYS82	Father's occupation
BYS83	Mother's and father's education

♦Variable not included in any release file.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

This page is intentionally blank.

Appendix D
Public-Use Masked/Suppressed Variables Available on Restricted Files for Licensed Users

This page is intentionally blank.

Appendix D
Public-Use Masked/Suppressed Variables Available on Restricted Files for Licensed Users

NOTE: The restricted-use files contain all variables on the public-use file. However, for purposes of protecting confidentiality, versions may differ in the amount of available detail (e.g., sometimes a given variable may appear in categorical form in the public-use file, but appear in continuous form in the restricted-use file, or it may include additional breakouts of collapsed categories, such as a restricted-use breakout for Native Hawaiians). In addition, a number of variables appear on the restricted file that have no counterpart on the public-use files (e.g., various geocode variables below the level of the four U.S. Census regions reported on the public-use file). The list below follows variable position order on the restricted-use Electronic Codebook (ECB).

Table D-1. Restricted-use unique variables, Education Longitudinal Study of 2002 (ELS:2002) base year student-level and school-level megafiles: 2002

Variable name	Variable description
Student-level variables	
BYEXPWT	Student expanded sample weight
RACE_R	Student's race/ethnicity—composite (restricted)
RACE2	Student's race/ethnicity—64 category
BYSARACE	Student's race/ethnicity—school roster
BYRACE_1	Student is White—composite
BYRACE_2	Student is Black or African American—composite
BYRACE_3	Student is Asian—composite
BYRACE_4	Student is Native Hawaiian/Pacific Islander—composite
BYRACE_5	Student is American Indian/Alaska Native—composite
HISPANIC	Student's Hispanic subgroup—composite
ASIAN	Student's Asian subgroup—composite
DOBIRTHR	Student's date of birth: year-month-day
PARACE_R	Parent's race/ethnicity—composite (restricted)
BYQXDATR	Date of base year student questionnaire administration
PISARFLG	Whether included in PISA[1] reading score concordance sample
BYIEPTYP	Federal disability category for base year IEPs[2]
BYACCTYP	Base year questionnaire/test accommodations
BYTXMTH	Math test theta T score
BYTXMTI1	Math theta T score—multiple imputation value 1 of 5
BYTXMTI2	Math theta T score—multiple imputation value 2 of 5
BYTXMTI3	Math theta T score—multiple imputation value 3 of 5
BYTXMTI4	Math theta T score—multiple imputation value 4 of 5
BYTXMTI5	Math theta T score—multiple imputation value 5 of 5
BYTXRTH	Reading test theta T score
BYTXRTI1	Reading theta T score—multiple imputation value 1 of 5
BYTXRTI2	Reading theta T score—multiple imputation value 2 of 5
BYTXRTI3	Reading theta T score—multiple imputation value 3 of 5
BYTXRTI4	Reading theta T score—multiple imputation value 4 of 5
BYTXRTI5	Reading theta T score—multiple imputation value 5 of 5

See notes at end of table.

Table D-1. Restricted-use unique variables, Education Longitudinal Study of 2002 (ELS:2002) base year student-level and school-level megafiles: 2002—Continued

Variable name	Variable description
BYRESZIP	Residential ZIP code for student/family
BYERAC_R	English teacher's race/ethnicity—composite (restricted)
BYSF1R_R	1st friend's race (restricted)
BYSF2R_R	2nd friend's race (restricted)
BYSF3R_R	3rd friend's race (restricted)
BYERAC_R	English teacher's race/ethnicity—composite (restricted)
BYMRAC_R	Math teacher's race/ethnicity—composite (restricted)
BYG10ER	Grade 10 enrollment—2001–02 school roster
BYCENDIV	Census division of school locale
BYSTATE	State code for school locale
BYCOUNTY	County code for school locale
BYSCHZIP	School ZIP code
HISPANIM	Imputation flag—Hispanic
ASIANIM	Imputation flag—Asian
BYS16	Student's Hispanic subgroup
BYS17A	Student is White
BYS17B	Student is Black/African American
BYS17C	Student is Asian
BYS17D	Student is Native Hawaiian/Pacific Islander
BYS17E	Student is American Indian/Alaska Native
BYS18	Student's Asian subgroup
BYS25CAA	1st friend is White
BYS25CAB	1st friend is Black/African American
BYS25CAC	1st friend is Asian
BYS25CAD	1st friend is Native Hawaiian/Pacific Islander
BYS25CAE	1st friend is American Indian/Alaska Native
BYS25CBA	2nd friend is White
BYS25CBB	2nd friend is Black/African American
BYS25CBC	2nd friend is Asian
BYS25CBD	2nd friend is Native Hawaiian/Pacific Islander
BYS25CBE	2nd friend is American Indian/Alaska Native
BYS25CCA	3rd friend is White
BYS25CCB	3rd friend is Black/African American
BYS25CCC	3rd friend is Asian
BYS25CCD	3rd friend is Native Hawaiian/Pacific Islander
BYS25CCE	3rd friend is American Indian/Alaska Native
BYS63	Occupation expects to have after high school—verbatim
BYS64	Occupation expects to have at age 30—verbatim
BYS68	Student's native language
BYS81A	Mother/female guardian's occupation—verbatim
BYS81B	Mother/female guardian's main job duties—verbatim
BYS82A	Father/male guardian's occupation—verbatim
BYS82B	Father/male guardian's main job duties—verbatim
BYP14	Parent's Hispanic subgroup
BYP15A	Parent is White
BYP15B	Parent is Black or African American
BYP15C	Parent is Asian
BYP15D	Parent is Native Hawaiian/Pacific Islander
BYP15E	Parent is American Indian/Alaska Native

See notes at end of table.

Table D-1. Restricted-use unique variables, Education Longitudinal Study of 2002 (ELS:2002) base year student-level and school-level megafiles: 2002—Continued

Variable name	Variable description
BYP16	Parent's Asian subgroup
BYP19A	Mother's occupation before coming to US
BYP19B	Mother's main job duties outside US
BYP22A	Father's occupation before coming to US
BYP22B	Father's job main duties outside US
BYP29	Native language of parent respondent
BYP39A	Parent's current/most recent job for pay in US
BYP39B	Parent's main job duties
BYP43A	Spouse/partner's current/most recent job for pay in US
BYP43B	Spouse/partner's main job duties
BYTE24A	Teacher is White (English)
BYTE24B	Teacher is Black/African American (English)
BYTE24C	Teacher is Asian (English)
BYTE24D	Teacher is Native Hawaiian/Pacific Islander (English)
BYTE24E	Teacher is American Indian/Alaska Native (English)
BYTM24A	Teacher is White (math)
BYTM24B	Teacher is Black/African American (math)
BYTM24C	Teacher is Asian (math)
BYTM24D	Teacher is Native Hawaiian/Pacific Islander (math)
BYTM24E	Teacher is American Indian/Alaska Native (math)
School-level variables	
BYSCMDST	Base year library media center questionnaire status
BYG10ER	Grade 10 enrollment—2001–02 school roster
BYCENDIV	Census division of school locale
BYSTATE	State code for school locale
BYCOUNTY	County code for school locale
BYSCHZIP	School ZIP code
BYNCESDI	NCES school district identification number
BYNCESSI	School identification number from CCD[3] or PSS[4]
BYA01	Total student enrollment as of October 2001
BYA02A	School has prekindergarten
BYA02B	School has kindergarten
BYA02C	School has 1st grade
BYA02D	School has 2nd grade
BYA02E	School has 3rd grade
BYA02F	School has 4th grade
BYA02G	School has 5th grade
BYA02H	School has 6th grade
BYA02I	School has 7th grade
BYA02J	School has 8th grade
BYA02K	School has 9th grade
BYA02L	School has 10th grade
BYA02M	School has 11th grade
BYA02N	School has 12th grade
BYA02O	School has 13th grade or higher
BYA03A	Comprehensive public school
BYA03B	Public magnet school

See notes at end of table.

Table D-1. Restricted-use unique variables, Education Longitudinal Study of 2002 (ELS:2002) base year student-level and school-level megafiles: 2002—Continued

Variable name	Variable description
BYA03C	Public magnet school with theme
BYA03D	Public school of choice
BYA03E	Year-round school
BYA03F	Area vocational school/center
BYA03G	Full-time technical/vocational school
BYA03H	Other technical or vocational school
BYA03I	Catholic diocesan school
BYA03J	Catholic parish school
BYA03K	Catholic religious order school
BYA03L	Catholic independent school
BYA03M	Other private school with religious affiliation
BYA03N	Private school without religious affiliation
BYA03O	Boarding school
BYA03P	Indian reservation school
BYA03Q	Military academy
BYA03R	Alternative/dropout prevention/continuation school
BYA03S	Charter school
BYA21	Percentage 10th graders receive free/reduced-price lunch
BYA22A	Number of full-time teachers

[1] PISA = Program for International Student Assessment.
[2] IEP = individualized education program.
[3] CCD = Common Core of Data.
[4] PSS = Private School Survey.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Appendix E
Glossary of Terms

This page is intentionally blank.

Appendix E
Glossary of Terms

Accommodations (testing): In ELS:2002, certain accommodations were offered to students with barriers to participation, such as students with disabilities or students with limited English proficiency. An accommodation is a change in how a test is presented, in how a test is administered, or in how the test taker is allowed to respond. This term generally refers to changes that do not substantially alter what the test measures. The proper use of accommodations does not substantially change academic level or performance criteria. Appropriate accommodations are made to provide equal opportunity to demonstrate knowledge. Examples of test accommodations include allowing extra time, use of a large-print version of a test, or conveying instructions in sign language. Cases in which accommodations were implemented in ELS:2002 are specially flagged (the indicator is BYTXACC).

Adaptive testing: In the ELS:2002 base year, multiple test forms of varying levels of difficulty were assigned based on the examinee's score on a routing test. Thus the specific sequence of questions that each student answered was tailored to that student's ability level. An advantage of adaptive tests is that reliability per unit of testing time is greater than in a nonadaptive test. Adaptive procedures help to minimize floor and ceiling effects (see "Ceiling Effect" and "Floor Effect"). ELS:2002 adaptive testing relies on Item Response Theory (see "IRT") assumptions to place students who have taken different test forms on the same vertical score scale. In the first follow-up, each student's test form will be assigned on the basis of base year test performance.

American Indian or Alaska Native: An American Indian or Alaska Native is a person who has origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.

Asian: An Asian is a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Base weights: See "Design Weights."

Bias: Bias is the difference between the reported value and the true value. Thus the bias of an estimate is the difference between the expected value of a sample estimate and the corresponding true value for the population. Response bias is the difference between respondent reports and their behavior or characteristics. Nonresponse bias is the difference that occurs when respondents differ as a group from nonrespondents on a characteristic being studied. Sample bias is the unequal selection or the omission of members of the population, without appropriate weighting. Relatedly, undercoverage bias arises because some portion of the potential sampling frame is missed or excluded, or there are duplicate units. For example, if the school list from which a school sample is drawn is incomplete or inaccurate (owing, for example, to the birth of new schools subsequent to the time the list was drawn up), school undercoverage may occur. (See "Nonresponse Bias" and "Bias Analysis.")

Bias analysis: Nonresponse bias analysis compares the characteristics of respondents and nonrespondents. Both unit nonresponse (school, student) and item nonresponse on questionnaires were subject to bias analyses in ELS:2002. For example, certain key data items were obtained for both responding and nonresponding schools, so that a school nonresponse analysis could be conducted, and bias in school-level estimates quantified.

Black or African American: A person having origins in any of the black racial groups of Africa.

Burden: Formally, burden is the aggregate hours realistically required for data providers to participate in a data collection. Burden also has a subjective or psychological dimension: the degree to which providing information is regarded as onerous may depend on the salience to the respondent of the questions that are being posed and on other factors, such as competing time demands.

Carnegie unit: A standard of measurement used for secondary education that represents the completion of a course that meets one period per day for 1 year.

CAPI: Computer-assisted personal interviewing, in which the questionnaire is loaded into a field interviewer's laptop computer.

CATI: Computer-assisted telephone interviewing.

CCD: Common Core of Data. Data annually collected from all public schools in the United States by NCES. Data from the CCD supplied the public school sampling frame for the ELS:2002 base year.

CD-ROM: ELS:2002 data are distributed primarily in an optical laser disc medium, specifically, CD-ROM (Compact Disc Read-Only Memory). A CD-ROM is a computer storage disc in the same physical form as an audio CD; it can store approximately 650 megabytes of digital data.

Ceiling effect: The result of a test having insufficient numbers of the more difficult items. In a longitudinal study, ceiling effects in the follow-up can cause change scores to be artificially constrained for high-ability examinees. The measurement problems related to floor and ceiling effects in combination with regression effects found at the extreme score ranges seriously hamper the accuracy of change measures in longitudinal studies. More information (i.e., smaller error of measurement) is obtained with respect to ability level if high-ability individuals receive relatively harder items (and if low-ability individuals receive proportionately easier items). The matching of item difficulty to a person's ability level yields increased reliability at the extremes of the score distribution, where it is most needed for studies of longitudinal change. A strategy employed in ELS:2002 to minimize ceiling (and floor) effects is to employ test forms that are "adaptive" to the ability level of the examinee. Multilevel tests—with second stage test assignment that is based on the first stage (routing test) performance work—minimize the possibility that ceiling effects might bias the estimates of the score gains. (See "Floor Effect" and "Adaptive Test.")

Classical test theory: Classical test theory postulates that a test score can be decomposed into two parts—a true score and an error component; that the error component is random with a mean of zero and is uncorrelated with true scores; and that true scores, observed scores, and error components are linearly related.

Closed-ended: A type of question in which the data provider's responses are limited to given alternatives (as opposed to an open-ended question. See "Open-ended").

Clustering: A sample selection method in which small geographical areas such as schools (as is the case in ELS:2002), school districts, counties, or residential blocks are selected as an initial stage, with individuals selected in a subsequent step. (See "Primary Sampling Unit.")

Cluster size: The number of ELS:2002 sample members attending a particular high school.

Codebook: Documentation of each variable being measured, including variable name, columns occupied by each variable in the data matrix, values used to define each variable, unweighted frequencies, unweighted percents, and weighted valid percents. (See "Electronic Codebook.")

Coefficient of variation: The ratio of the standard deviation of an estimate to the value of the estimate.

Cognitive test battery: One of the two parts of the student survey (the second part being the student questionnaire). Two achievement areas (mathematics and reading) were measured in the base year.

Cohort: 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 will also contain, however, a nationally representative sample of high school seniors in the spring term of the 2003–04 school year (see "Freshening"). In contrast, the Program for International Student Assessment (PISA) is an age cohort, based on students who were 15.25 years of age in April of 2000 or 2003.

Composite variable: A composite variable is one that is constructed through either 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 that is calculated through the application of a mathematical function or transformation to a variable (e.g., conversion of raw test scores to percentile ranks). Also called a "derived variable," "created variable," or "constructed variable."

Confidence interval: A sample-based estimate expressed as an interval or range of values within which the true population value is expected to be located (with a specified degree of confidence).

Confidentiality protections: NCES is required by law to protect individually identifiable data from unauthorized disclosure. To this end, the ELS:2002 data have been subject to a disclosure risk analysis to determine which records require masking to produce the public-use data file from the restricted-use data file. Disclosure coarsening techniques (such as recoding of continuous variables into categorical, top and bottom coding, and so on), as well as data perturbation

techniques (e.g., data swapping) have been used to provide disclosure protection to the ELS:2002 data. (See also "Data Swapping" and "Disclosure Risk Analysis.")

Consent, active (explicit): One variety of informed consent is called active or explicit consent. Typically, in active consent, a signed agreement to participate in a study must be obtained. In ELS:2002, permission of parents was required before students could be surveyed. Some schools required active parental consent (i.e., that a signed permission form be obtained).

Consent, passive (implied): Another variety of informed consent is called passive or implied consent. In passive consent, a permission form is sent to the relevant party (in ELS:2002, normally the parent or guardian of the sampled student), who has the opportunity to return the form to indicate denial of permission. If the form is not returned, it is assumed that the individual has no objection to survey participation. In ELS:2002, most schools allowed passive parental consent for their sophomore's participation in the study.

Constructed response item: In the ELS:2002 assessment battery, a non-multiple-choice item that requires some type of written response.

Contextual data: In ELS:2002, the primary unit of analysis is the student, and information from the other study components, referred to as contextual data, should be viewed as extensions of the student data. For example, observations made in school administrator, teacher, librarian, and parent reports on the student's school learning environment or home situation would be considered contextual data.

Coverage rate: In ELS:2002 base year contextual samples, 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). (See "Response Rate.") 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 on the sampling frame (undercoverage), or duplicated or erroneously enumerated units (overcoverage) (see "Bias" for discussion of undercoverage bias).

Cross-sectional analysis: A cross-sectional design represents events and statuses at a single point in time. For example, a cross-sectional survey may measure the cumulative educational attainment (achievements, attitudes, statuses) of students at a particular stage of schooling, such as 10th or 12th grade. In contrast, a longitudinal survey (or repeated measurement of the same sample units) measures the change or growth in educational attainments that occurs over a particular period of schooling. The longitudinal design of ELS:2002 generates two representative cross-sections (high school sophomores in 2002, and, through sample freshening, seniors in 2004). It also permits analysis of individual-level change over time through longitudinal analysis and of group-level and intercohort change through the cross-sectional

comparisons to past studies of similarly defined grade cohorts. (See "Longitudinal or Panel Survey" and "Cross-cohort Analysis.")

Cross-cohort (or intercohort) analysis: The ELS:2002 base year survey contained many data elements that were comparable to items from prior studies. They will supply a basis for comparison with earlier sophomore cohorts (such as 1980 sophomores in the High School and Beyond [HS&B] longitudinal study and 1990 sophomores in the National Education Longitudinal Study of 1988 [NELS:88]). With a freshened senior sample, the ELS:2002 first follow-up will support comparisons to 1972 (National Longitudinal Study of the High School Class of 1972 [NLS-72]), 1980 (HS&B), 1992 (NELS:88), and 2004 high school seniors. The first follow-up academic transcript component will offer a further opportunity for cross-cohort comparisons with the high school transcript studies of HS&B, NELS:88, and the National Assessment of Educational Progress (NAEP). With three or more timepoints, *trend analyses* are possible. With ELS:2002, this condition has now been met for the sophomore cohorts; trend studies of the senior cohorts were initiated with NELS:88. Essentially, three kinds of intercohort comparison are possible. First, cohorts can be compared on an *intergenerational or cross-cohort time-lag basis*. Both cross-sectional and longitudinal time-lag comparisons may be made. An example of a cross-sectional time-lag comparison would be looking at the status of HS&B (1980), NELS:88 (1990), and ELS:2002 (2002) sophomores to see how the situation of sophomores has changed over time. An example of longitudinal time-lag comparison would be an examination of the magnitude and correlates of achievement gain of HS&B, NELS:88, and ELS:2002 sophomores over the last 2 years of high school. Second, *fixed-time comparisons* are also possible, in which groups within each study are compared at different ages but the same point in time (e.g., the NLS-72, HS&B senior, and HS&B sophomore cohorts all could be looked at in 1986, some 14, 6, and 4 years after each respective cohort graduated from high school). Such a perspective would permit one to compare, for example, employment rates for 22-, 24-, and 32-year-old high school graduates). Finally, *longitudinal comparative analysis* of the cohorts can be performed by modeling the history of the grade cohorts.

Cut score: A cut score is a specified point on a score scale such that scores at or above that point are interpreted or acted upon differently from scores below that point.

Data element: The most basic unit of information. In data processing, it is the fundamental data structure. It is defined by its size (in characters) and data type (e.g., alphanumeric, numeric only, true/false, date) and may include a specific set of values or range of values.

Data swapping: Data swapping is defined in the *NCES Statistical Standards* as a perturbation disclosure limitation technique that results in a confidentiality edit. An example of data swapping would be to assume a data file has two potential individual identifying variables, for example, sex and age. If a sample case needs disclosure protection, it is paired with another sampled case so that each element of the pair has the same age, but different sexes. The data on these two records are then swapped. After the swapping, anyone thinking they have identified either one of the paired cases gets the data of the other case, so they have not made an accurate match and the data have been protected. (See also "Confidentiality Protections.")

Design effect: A measure of sample efficiency. The design effect (DEFF) is the variance of an estimate divided by the variance of the estimate that would have occurred if a sample of the same

size had been selected using simple random sampling. Sometimes it is more useful to work with standard errors than with variances. The root design effect (DEFT) expresses the relation between the actual standard error of an estimate and the standard error of the corresponding estimates from a simple random sample. (See also "Effective Sample Size.")

Design weights: Design weights compensate for unequal probabilities of selection. More specifically, the design weight is the inverse of the probability of selection. Design weights are also called raw weights, base weights, unadjusted weights, or sampling weights. Design weights may be contrasted to adjusted weights (adjusted to compensate for nonresponse, and also called final weights or analysis weights). Roughly, the design weight is calculated as the inverse of the probability of selection, taking into account all stages of the sample selection process. More precisely, design weights are the inverses of the expected frequencies with which population units appear in conceptually repeated samples selected using the sampling design developed for the study. Unlike the final weights, design weights are generated for all sample members, respondents and nonrespondents alike. Design weights do not appear on the ELS:2002 public-use files. (See also "Final Weights" and "Sampling Weights.")

Differential Item Functioning (DIF)**:** DIF exists when examinees of equal ability differ on an item solely because of their membership in a particular group (e.g., if an item favors males over females, or one racial or ethnic group over another, and cannot be explained by relevant factors such as differential coursetaking). DIF for ELS:2002 items was examined in the base year field test and is reported in the ELS:2002 Base Year Field Test Report. Items with DIF problems were revised or deleted.

Disability: A disability is a physical or mental impairment that substantially limits one or more of the major life activities (Title 42 U.S.C. Section 12102).

Disclosure risk analysis: Investigation of study data to evaluate and minimize the risk of identification of individual sample units, to preserve the confidentiality of the data. ELS:2002 data have been subjected to a disclosure risk analysis to protect confidential information about individual respondents; see the entry for "Public-use Data File." For a more detailed account of disclosure risk analysis, and of means of altering data (including masking, data perturbation, and data swapping) to prevent disclosure, see the current NCES Statistical Standards document.

Domain: A domain refers to a defined universe of knowledge, skills, abilities, attitudes, interests, or other human characteristics.

Effective sample size: Effective sample size may be defined as the ratio of the raw sample size divided by the design effect. (For example, the sampling variance of a mean standard score is equal to the reciprocal of the effective sample size, not the reciprocal of the raw sample size.) In essence, then, effective sample size is the sample size under a simple random sample design that is equivalent to the actual sample under the complex sample design, wherein the actual sample size is determined by multiplying the effective sample size by the anticipated design effect. (See also "Design Effect.")

Electronic codebook (ECB): While hardcopy codebooks with item stems, response categories, associated response frequency distributions, unweighted percents, and weighted valid percents

are contained within the ELS:2002 base year user's manual, ELS:2002 data are also available on CD-ROM in an electronic codebook (ECB) format. Electronic codebooks are menu-driven systems that allow users to perform functions such as the following: (a) search a list of database variables based upon key words or variable names/labels, (b) display unweighted percentages for each variable in the database, (c) display question text for each variable in the database, (d) select or tag variables for subsequent analysis, (e) generate SAS-PC or SPSS-PC+ program code/command statements for subsequently constructing a system file of the selected variables, and (f) generate a codebook of the selected variables.

Equating: Equating of two tests is established when examinees of every ability level and from every population group can be indifferent about which of two tests they take. Not only should they have the same expected mean score on each test, but they should also have the same errors of measurement. In contrast, test *linkage* results from placing two or more tests on the same scale, so that scores can be used interchangeably. (See also "Equated Test Score.")

Equated test score: Test equating takes place in two distinct contexts in ELS:2002. One context is *vertical equating* of forms for use in successive grades, such that the achievement growth of individual ELS:2002 sample members over time can be accurately measured. Another context is *cross-sectional equating* and *linking*, as to other tests (e.g., the National Education Longitudinal Study of 1988 [NELS:88], the Program for International Student Assessment [PISA], and the National Assessment of Educational Progress [NAEP]).

ETS: Educational Testing Service. RTI's subcontractor for ELS:2002 cognitive test development, scoring, and scaling.

Expanded sample: Although no sophomores were excluded from ELS:2002, those who could not validly be assessed or who could not validly complete the student questionnaire(e.g., students with a severe disability or limitation in their knowledge of the English language), were not eligible for these components. Contextual data (parent, teacher, school administrator) reports were collected for this group. In the first follow-up, their transcripts will be collected and the eligibility status of each will be re-evaluated. The expanded sample comprises all ELS:2002 sophomores; that is, both those who were eligible to complete the student questionnaire and test, and those who were not.

Facilities checklist: Completed by the RTI survey administrator, the facilities checklist is designed to extend the information available about the school by providing data on the school buildings and grounds that will help researchers to understand the adequacy and appearance of the school's physical plant, its safety and security features, and its role as a constituent of the school's general environment.

File: Refers to a data file containing a set of related computerized records.

Final weights: Final weights are sometimes called nonresponse-adjusted weights, adjusted weights, or analysis weights. Building on the design (raw) weight, they compensate for nonresponse. (See "Design Weights.")

Floor effect: The result of a cognitive test being too difficult for a large number of the examinees, causing the low-ability examinees to receive chance scores on the first testing, and on

subsequent testings if the test remains too difficult. Floor effects result in an inability to discriminate among low-ability individuals at time one or time two, and there will be no reliable discrimination among examinees with respect to amounts of change. A possible solution, utilized in ELS:2002, is to develop test forms that are "adaptive" to the ability level of the examinee, which tends to minimize the possibility of floor effects biasing the estimates of the score gains. (See also "Ceiling Effect" and "Adaptive Testing.")

Frame: A list of all the sampling units that represent the population. The Common Core of Data (CCD) and Private School Survey (PSS) were drawn upon for the ELS:2002 school frame. For an implicit list of the nation's high school sophomores as of spring term 2002, school rosters from participating schools listing their sophomore class were relied on.

Frame population: The set of elements (e.g., schools) that can be enumerated prior to the selection of a survey sample.

Freshening: A freshened sample includes cases from the longitudinal sample of a data set, plus new cases added to produce cross-sectional estimates of the population at the time of a subsequent wave of a longitudinal data collection. In the National Education Longitudinal Study of 1988 (NELS:88), freshening was the means by which high school sophomores were added in the first follow-up who were not in the eighth grade in the United States 2 years before. This process was repeated in the second follow-up, adding high school seniors who were not in the eighth grade in the United States 4 years before, and not in the tenth grade in the United States 2 years before. This process ensured that the sample would be representative of the 1992 senior class by allowing 1992 seniors who did not have a chance for selection into the base year (or the first follow-up) sample to have some probability of 1992 selection. The same procedure will be implemented in ELS:2002 in the 2004 first follow-up to ensure a nationally representative senior cohort. (See also "Half-open Interval.")

Half-open interval: A technique used to increase coverage. It is usually applied to a new list that includes cases that were covered in a previous frame, as well as new in-scope units not included in the previous frame. In this technique, new in-scope units between unit A on the previous frame up to, but not including, unit B (the next unit on the previous frame) are associated with unit A. These new units have the same selection probability as do unit As. This process is repeated for every unit on the previous frame. The new units associated with the actual sample cases are now included in the sample with their respective selection probabilities (freshening). Student sample freshening in the National Education Longitudinal Study of 1988 (NELS:88) first and second follow-ups, and the freshening to be conducted in the ELS:2002 first follow-up, rely on such a procedure. The half-open interval procedure was also used for ELS:2002 base year sample updating prior to Survey Day. (See also "Freshening" and "Sample Updating or *Refreshing*.")

Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race. The term "Spanish origin" can be used in addition to "Hispanic or Latino."

HS&B (High School and Beyond). The second in the series of longitudinal high school cohort studies sponsored by NCES. The HS&B base year study surveyed sophomore and senior

students in 1980. The sophomore cohort was last interviewed in 1992 and their postsecondary transcripts collected in 1993. The senior cohort was last interviewed in 1986.

Imputation: Imputation involves substituting values for missing or inconsistent data in a data set. Prediction of a missing value is typically based on a procedure that uses a mathematical model in combination with available information. Missing data for key items in ELS:2002 have been imputed.

Individualized education program (IEP): A written statement or plan for each individual with a disability that is developed, reviewed, and revised in accordance with Title 42 U.S.C. Section 1414(d).

Individually identifiable data: Data from any record, response form, completed survey, or aggregation about an individual or individuals from which information about particular individuals may be revealed.

Instrument: An evaluative device that includes tests, scales, and inventories to measure a domain using standardized procedures.

IRT: Item Response Theory. A method of estimating achievement level by considering 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. Another attribute of IRT that makes it useful for ELS:2002 is the calibration of item parameters for all 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 will also permit vertical scaling of the two grade levels (10th grade in 2002, 12th grade in 2004). (See, in contrast, "Classical Test Theory.")

Item nonresponse: The amount of missing information when a valid response to an item or variable was expected. (See "Unit Nonresponse" and see "Bias Analysis.")

LEP: Limited English proficient. A concept developed to assist in identifying those language-minority students (individuals from non-English language backgrounds) who need language assistance services, in their own language or in English, in the schools. (See "NEP" and "LM.") A limited English proficient student is one who meets one or more of the following conditions:

a. the student was born outside of the United States or the student's native language is not English,

b. the student comes from an environment in which a language other than English is dominant, or

c. the student is an American Indian or Alaska Native and comes from an environment in which a language other than English has had a significant impact on his/her level of English language proficiency,

and who has such difficulty speaking, reading, writing, or understanding the English language as to deny him or her the opportunity to learn successfully in English-only classrooms.

LM: Language Minority. A non-, limited-, or fully English-proficient student in whose home a non-English language is typically spoken.

Library media center questionnaire: This instrument supplies information about library/media center organization and staffing, technology resources, extent of library and media holdings, student access to and use of the library/media center, and its role in supporting the school's curriculum.

Longitudinal or panel survey: In a longitudinal design, similar measurements—of the same sample of individuals, institutions, households, or of some other defined unit—are taken at multiple time points. ELS:2002 employs a longitudinal design that follows the same individuals over time and permits the analysis of individual-level change. (See "Cross-sectional Survey.")

Machine editing: Also called forced data cleaning or logical editing. Uses computerized instructions (including logical or deductive imputation) in the data cleaning program that ensure common-sense consistency within and across the responses from a data provider.

Microdata (microrecords): Observations of individual sample members, such as those contained on the ELS:2002 data files.

MPR Associates: An RTI subcontractor for the ELS:2002 base year and first follow-up studies.

NAEP: The National Assessment of Educational Progress. NAEP is a cross-sectional assessment program that measures achievement at the group level for students in fourth, eighth, and twelfth grades and provides a time series for measuring trends in academic progress of 9-, 13-, and 17-year olds. ELS:2002 tests differ from but complement those of NAEP by providing a basis for measuring individual-level achievement growth between 10th and 12th grades in mathematics and relating cognitive gains in this subject to the individual, school, and family factors and processes that are measured in the various ELS:2002 questionnaires and school records (transcript) studies.

Native Hawaiian or Other Pacific Islander: Any person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

NCES: The National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. This governmental agency is the sponsor of ELS:2002 and is also the sponsoring agency for (among other studies) the National Assessment of Educational Progress (NAEP), the National Education Longitudinal Study of 1988 (NELS:88), the High School and Beyond (HS&B) longitudinal study, and the National Longitudinal Study of the High School Class of 1972 (NLS-72).

NELS:88: The National Education Longitudinal Study of 1988. Third in the series of longitudinal high school cohort studies sponsored by NCES. The study represents three cohorts: the eighth-grade class of 1988, the sophomore class of 1990, and the senior class of 1992. The study collected questionnaire and test data in 1988, 1990, and 1992 on students' school experiences, as well as background information from school administrators, teachers, parents (in the base year and second follow-up only), and school records. Data on postsecondary and out-of-school experiences were collected in interviews conducted in 1994 and 2000 and through a postsecondary education transcripts study in 2000–01.

NEP: No English proficiency. A student who does not speak English. (See "LEP.")

Noncoverage: Units of the target population that are missing from the frame population. Includes the problems of incomplete frames and missing units.

Nonresponse: (See "Item Nonresponse," "Unit Nonresponse," "Bias Analysis," and "Nonresponse Bias.")

Nonresponse bias: Nonresponse bias may occur as a result of not obtaining 100 percent response from the selected cases. More specifically, nonresponse bias occurs when the expected observed value deviates from the population parameter. The potential magnitude of nonresponse bias is estimated as the product of the nonresponse rate and the difference in values of a characteristic between respondents and nonrespondents. (See also "Bias" and "Bias Analysis.")

NLS-72: The National Longitudinal Study of the High School Class of 1972. This project was the first in the series of longitudinal high school cohort studies sponsored by NCES. The final round of data collection took place in 1986.

Nonsampling error: An error in sample estimates that cannot be attributed to sampling fluctuations. Such errors may arise from many sources, including imperfect implementation of sampling procedures, differential unit or item nonresponse across subgroups, bias in estimation, or errors in observation and recording.

OMB: The Office of Management and Budget, U.S. Executive Branch. OMB is a federal agency with the responsibility for reviewing all studies funded by executive branch agencies. OMB reviewed, commented on, and approved the ELS:2002 questionnaires, as indicated by their approval number and its expiration date in the top right corner of the questionnaire covers.

Open-ended: A type of question in which the data provider's responses are not limited to given alternatives.

Optical disc: A disc that is read optically (e.g., by laser technology), rather than magnetically. (See "CD-ROM.")

Optical scanning: A system of recording responses that transfers responses into machine-readable data through optical mark reading. This method of data capture was used for the ELS:2002 student questionnaires and cognitive tests, as well as for the school administrator, teacher, and library media center questionnaires, and hardcopy (as contrasted to CATI [computer-assisted telephone interviewing]) administrations of the parent questionnaire.

Oversampling: Deliberately sampling a portion of the population at a higher rate than the remainder of the population. For example, in ELS:2002, private schools have been oversampled. Within schools, Asians have been oversampled.

Parent/guardian questionnaire: The ELS:2002 parent component sought to collect information from parents of all base year student sample members. The parent or guardian who knew most about his or her child's educational experience was asked to complete the questionnaire.

PISA: The Program for International Student Assessment assesses 15-year-olds in reading, mathematics, and science. In 2000, the primary focus of the assessment was reading. The United States and 31 other nations participated, under the aegis of the Organization for Economic Cooperation and Development (OECD). In 2003, the primary focus was mathematics, and in 2006, the primary focus will be science. A crosswalk (or concordance) has been developed between the ELS:2002 reading test and the PISA reading test, so that the PISA scale can be implemented in ELS:2002. A similar scale linkage will be effected between the ELS:2002 mathematics test (2002) and the PISA math test (2003).

Population: All individuals in the group to which conclusions from a data collection activity are to be applied. Weighted results of ELS:2002 data provide estimates for populations and subgroups.

Population variance: A measure of dispersion defined as the average of the squared deviations between the observed values of the elements of a population or sample and the population mean of those values.

Postsecondary education: The provision of formal instructional programs with a curriculum designed primarily for students who have completed the requirements for a high school diploma or equivalent. This includes programs of an academic, vocational, and continuing professional education purpose, and excludes vocational and adult basic education programs.

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 in terms of the sampling error (or standard error) of an estimate.

Primary sampling unit (PSU): Unit chosen at the first stage of a cluster sample. In ELS:2002, the PSU is the school; in other studies, geographical units such as a county or metropolitan statistical area (MSA) may serve as the PSU.

Probability sample: A sample selected by a method such that each unit has a fixed and determined probability of selection—i.e., each population unit has a known, nonzero chance of being included.

Proficiency score: Proficiency scores (or criterion-referenced mastery scores) are based on clusters of items within each test that are of similar content and difficulty. Both normative (e.g., achievement quartiles) and proficiency scores are available from the ELS:2002 database.

PSS: Private School Survey. An NCES universe survey encompassing the nation's private schools. PSS was the private school sampling frame for the ELS:2002 base year.

Public-use data file: A public-use file includes a subset of data that have been coded, aggregated, or otherwise altered to mask individually identifiable information; it thus is available to all external users. Unique identifiers, geographic detail, and other variables that cannot be suitably altered are not included in public-use data files. Public-use edits are based on an assumption that external users have access to both individual respondent records and secondary data sources that include data that could be used to identify respondents. For this reason, the editing process is relatively extensive. When determining an appropriate masking process, the public-use edit takes into account and guards against matches on common variables from all known files that could be matched to the public-use file. The analysis used to determine which records require masking is called a disclosure risk analysis**.**

Range check: A determination of whether responses fall within a predetermined set of acceptable values.

Record format: The layout of the information contained in a data record (includes the name, type, and size of each field in the record).

Records: A logical grouping of data elements within a file upon which a computer program acts.

Refreshed student: See "Sample Updating or *Refreshing*."

Relative bias. Relative bias is the bias of the estimate divided by the estimate. It provides an indication of the order of magnitude of the bias with respect to the estimate.

Reliability: The consistency in results of a test or measurement including the tendency of the test or measurement to produce the same results when applied twice to some entity or attribute believed not to have changed in the interval between measurements.

Reserve code (or **reserved code**)**:** Certain codes have been reserved to stand for a number of situations in which missing data occurs in response frequencies. In ELS:2002, the reserve code conventions are as follows: -1 = "Don't know;" -2 = "Refused;" -3 = "Legitimate Skip/NA;" -4= Nonrespondent;" -5 = "Out of Range;" -6 = "Multiple Response;" -7 = "Not Administered—abbreviated interview;" and -9= "Missing."

Response rate: In general, unit response rates are calculated as the ratio of the weighted number of completed instruments to the weighted number of in-scope cases, using the sample base weight (the inverse of the probability of selection). In multistage samples, such as the base year of ELS:2002, overall response is the product of both stages (though for many purposes, the stages are reported separately). Item response rates are calculated as the ratio of the number of respondents for whom an in-scope response was obtained to the number of respondents who are asked to answer a given item. Calculation of unit and item response rates can be a complex matter, and additional considerations arise in reporting in follow-up waves of longitudinal studies, for composite (constructed) variables, and for other cases. More detailed information can be found by consulting NCES Standard 1-3 in the NCES 2002 Statistical Standards

document (available on the web at <http://nces.ed.gov/pubsearch/wnew.asp?1>). Bias analyses conducted when response rates are below targets help to assess any possible limitations to the generalizability of survey estimates. (See "Bias Analysis.")

Restricted-use data file: A restricted-use file includes individually identifiable information that is confidential and protected by law. Restricted-use data files are not required to include variables that have undergone public-use edits. ELS:2002 data are available in both public-use and restricted-use forms. Use of the restricted data requires the researcher to obtain a special license from NCES.

RTI International (RTI): A nonprofit university-affiliated research organization with headquarters at Research Triangle Park, North Carolina, that conducted the base year of ELS:2002 and is currently conducting the first follow-up of the study on behalf of NCES. RTI International is a trade name of Research Triangle Institute.

Sample: Subgroup selected, by a probability method, from the entire population, in order to represent it.

Sample updating or *refreshing***:** Because students can transfer into or out of a school after sampling, the base year student sample in ELS:2002 (as in High School and Beyond [HS&B] and the National Education Longitudinal Study of 1988 [NELS:88]) was updated to remove students who had transferred out and to give sophomores who had transferred in since sampling a chance of selection. The half-open interval procedure was employed for sample updating prior to Survey Day, using the school 10th-grade enrollment lists.

Sampling error: The part of the difference between a value for an entire population and an estimate of that value derived from a probability sample that results from observing only a sample of values.

Sampling frame. See "Frame" or "Frame population."

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: A multiplicative factor equal to the reciprocal of the probability of a respondent being selected for the study, with adjustment for nonresponse. The sum of the weights provides an estimate of the number of persons in the population represented by a respondent in the sample.

Scaling: Scaling refers to the process of assigning a scale score based on the pattern of responses. (See also "Equated Test Score" and "IRT.")

School administrator questionnaire: This questionnaire was to be completed by the base year principal and/or someone designated by the principal. The questionnaire sought basic information about school policies, number of students in each class, curriculum offered, programs for disadvantaged and disabled students, and other school characteristics.

School climate: The social system and ethos or culture of the school, including the organizational structure of the school and values and expectations within it.

School coordinator: A person designated in each school to act as a contact person between the school and RTI. This person assisted with establishing a Survey Day in the school and preparing for the survey.

Section 504: Section 504 of the Rehabilitation Act of 1973, as amended (Title 29 U.S.C. 794 Section 504), prohibits discrimination on the basis of handicap in federally assisted programs and activities.

Selection probability: The chance that a particular sampling unit has of being selected in the sample.

Simple random sampling (SRS): SRS uses equal probability sampling with no strata or clusters. The ELS:2002 sample is stratified and clustered. Most statistical analysis software assumes SRS and independently distributed errors. For studies such as ELS:2002, special variance estimation software (such as SUDAAN, WesVar, AM, or Stata) is required to compute the standard error of estimates.

Standard deviation: The most widely used measure of dispersion of a frequency distribution. 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.

Statistical significance: The finding (based on a derived probability, rather than a certitude) that two or more estimates are truly different from one another and not a merely apparent difference reflecting chance variation.

Stratification: The division of a population into parts, or strata. In a stratified sample, the total population is divided into strata or subgroups. Strata are created by partitioning the frame and are generally defined to include relatively homogeneous units within strata. Stratification is used to reduce sampling error. In ELS:2002, the sampling frame was sorted to create strata or subgroups of schools, and schools were selected independently within each stratum. Schools were stratified by superstrata (combinations of school type or sector and geographic region) and substrata (urban, suburban, rural).

Student questionnaire: One of the two parts of the ELS:2002 base year student survey (the other part is the cognitive test battery). This instrument contained a locator section for tracing sample members for future waves of ELS:2002 and a series of questions about school and home environments, time use, attitudes, values, and aspirations.

Survey Administrator: A member of RTI's field staff in charge of conducting in-school data collection sessions (see "Survey Day" below). The individual in this role was called a Team Leader in NELS:88 and a Survey Representative in HS&B.

Survey Day: A day chosen by the school during the data collection period when an RTI survey administrator and assistant administered the survey to the school's sample of students. The Survey Day session lasted about 2 hours. Two Make-up Days were normally offered for students who missed Survey Day.

Target population: The finite set of observable or measurable elements that will be studied, or the conceptual population of analytic units for which data are collected and estimates are made. In the ELS:2002 base year, the target population was spring term 2002 sophomores in all regular public and private schools with 10th grades in the 50 states and the District of Columbia.

Teacher questionnaire: In the base year, math and reading teachers of ELS:2002 sophomore participants were asked to complete a teacher questionnaire, which collected data on school and teacher characteristics (including teacher qualifications and experience) and evaluations of student performance.

Teacher sample: In the ELS:2002 base year, two teacher reports were sought for each student, one from the student's mathematics teacher and one from the student's English teacher.

Technical Review Panel (TRP): A TRP is a specially appointed, independent group of substantive, methodological, and technical experts who offer advice to the study's contractor on issues of study design and content. TRP members are nominated by the contractor and approved by NCES. Typically TRPs are convened at least once a year within the life of a contract.

Trimming: A process by which extreme weights are reduced (trimmed) to diminish the effect of extreme values on estimates and estimated variances.

Unit nonresponse: Failure of a survey unit (e.g., at the institutional level, a school, or at the individual level, a respondent, such as a student or a teacher) to cooperate or complete a survey instrument. *Overall unit nonresponse* reflects a combination of unit nonresponse across two or more levels of data collection, where participation at the second stage of data collection is conditional upon participation in the first stage of data collection. In ELS:2002, overall nonresponse is the product of school-level nonresponse times student nonresponse. *Total nonresponse* reflects a combination of the overall unit nonresponse and item nonresponse. (See also "Item Nonresponse" and "Nonresponse Bias.")

Urbanicity (or Metropolitan Status): The ELS:2002 school sample was stratified by metropolitan status or urbanicity, in accordance with the following three locale codes: (1) Urban: the school is in a large or mid-size central city; (2) Suburban: the school is in a large or small town or is on the urban fringe of a large or mid-size city; and (3) Rural: the school is in a rural area. Locale indicators were taken from the Common Core of Data (CCD) for public schools and the Private School Survey (PSS) for private schools.

Validity: The capacity of an item or instrument to measure what it was designed to measure, stated most often in terms of the correlation between scores in the instrument and measures of performance on some external criterion. It is the extent to which a test or set of operations measures what it is supposed to measure. Reliability, on the other hand, refers to consistency of measurement over time. (See "Reliability.")

Variance: The average of the squared deviations of a random variable from the expected value of the variable. The variance of an estimate is the squared standard error of the estimate. **(**See "Population Variance" and "Sampling Variance.")

Wave: A wave is a round of data collection in a longitudinal survey (e.g., the base year and each successive follow-up are each waves of data collection).

Weighted response rates: Unit response rates are calculated as the ratio of the weighted number of completed interviews to the weighted number of in-scope sample cases. Unit response rates are calculated using the sample base weights (inverse of the probability of selection).

Weighted estimates: Weighted estimates (as in the ELS:2002 codebook) are survey estimates in which the sample data are statistically weighted (multiplied) by factors reflecting the sample design. The general purpose of weighting is to compensate for unequal probabilities of selection into the sample and to adjust for the fact that not all schools or individuals selected into the sample actually participated. The design weights (also known as base weights, and typically equal to the reciprocals of the overall selection probabilities) are multiplied by a nonresponse or poststratification adjustment for a final weight. Thus, for example, in ELS:2002, the 752 participating schools in the base year represent a national population of 24,795 schools. Individual schools may "represent" anywhere from a minimum of one school to a maximum of 96 schools. To take a National Education Longitudinal Study of 1988 (NELS:88) example of weighted estimates, 12,111 base year questionnaire respondents reported themselves to be male, and a slightly greater number (12,244) reported themselves to be female. When these cases are multiplied by the nonresponse-adjusted student weights to yield a weighted percent that reflects the national population of eighth graders, the estimate for males is 50.1 percent of the 1988 eighth-grade cohort, while females are estimated to comprise 49.9 percent of the nation's 1988 eighth graders.

White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

This page intentionally blank.

Appendix F
Documentation for Imputed Variables

This page is intentionally blank.

Appendix F
Documentation for Imputed Variables

Table F–1. Imputed Education Longitudinal Study of 2002 (ELS:2002) base year variables

Imputation variables
Student sex
Student race/ethnicity
Student Hispanic subgroup
Student Asian subgroup
Father's educational attainment (composite)
Mother's educational attainment (composite)
Father's occupation/occupational prestige♦
Mother's occupation/occupational prestige♦
Family income (parent report)
Highest parental education (composite)
English as native language
Student IRT number-right score in reading
Student IRT number-right score in mathematics
Standardized T-score in reading
Standardized T-score in mathematics
Standardized T-score, reading + mathematics composite
Achievement quartile in reading
Achievement quartile in mathematics
Composite achievement quartile (reading + mathematics)
Probability of proficiency, reading (3 levels/variables)
Probability of proficiency, math (5 levels/variables)
Parent-reported family composition
Student educational aspirations
Parental aspirations for student
School region
School type (public, Catholic, other private)
School metropolitan type (urban, suburban, rural)
School percent minority
School grade 10 membership

♦Composite variable based on parent reports. When parent reports were missing, the variable was based on student reports and, if still missing, on imputation. Imputed for use in construction of the socioeconomic status variable. Not available on the data file.

NOTE: The presentation of imputation variables in Table F-1 differs slightly from the presentation in Table 23. Table 23 lists the ability estimate (*theta)* in mathematics and reading from which test variables were derived. Table F-1 lists the test variables derived from the imputed version of *theta*. While only *theta* was directly imputed, the imputed *theta* provided the basis for complete information about test performance for the scores listed in Table F-1. Also, several school-level variables, listed in F-1, were identified as key, and to be imputed if there were missing data for them. For these school-level imputation variables (region, school type, metropolitan status, percent minority, and grade 10 enrollment), in the event, 100 percent coverage was obtained from universe files, and statistical imputation was not required.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table F–2. Imputation classes and sort variables for variables imputed by weighted sequential hot deck imputation

Imputation variable	Imputation class variables	Sort variables
English as native language (STLANG)	Student race (RACE) Mother's birthplace (BYP17)	Census region (BYREGION) Urbanicity (BYURBAN) Percent minority Parent's race (PARACE) Father's birthplace (BYP20) Student birthplace (BYP23)
Student Hispanic origin (HISPANIC)	Friend race composite English as native language (BYS67)	School type (BYSCTRL) Census region (BYREGION) Urbanicity (BYURBAN) School (SCH_ID) Parent's race (PARACE)
Student Asian origin (ASIAN)	Friend race composite English as native language (BYS67)	School type (BYSCTRL) Census region (BYREGION) Urbanicity (BYURBAN) School (SCH_ID) Parent's race (PARACE)
Type of school program (SCHPROG)	School coed status (BYA11) Percent 10th graders in general HS (BYA14A) Percent 10th graders in college prep (BY14B) Percent 10th graders in voc/tech (BYA14D)	School type (BYSCTRL) Census region (BYREGION) Urbanicity (BYURBAN) Percent minority Percent 10th graders in LEP (BYA20)
Student educational expectations (STEXPECT)	Student sex (SEX) Type of school student wants to attend after HS (BYS58) Good grade importance (BYS37)	School type (BYSCTRL) Census region (BYREGION) Urbanicity (BYURBAN) Student race (RACE) School program (BYS26) How far parent expects student to go in school (BYP81) Parent saved money for 10th graders Education after high school (BYP82)

See notes at end of table.

Table F–2. Imputation classes and sort variables for variables imputed by weighted sequential hot deck imputation—Continued

Imputation variable	Imputation class variables	Sort variables
Parental aspirations for student postsecondary achievement (PARASPIR)	Student race (RACE) Student educational expectations (BYS56)	School type (BYSCTRL) Census region (BYREGION) Urbanicity (BYURBAN) Student race (RACE) School program (BYS26) Parental aspirations for student postsecondary achievement (BYP79) Parental savings for student postsecondary schooling (BYP82)
Family composition (BYFCOMP)	Parent marital status Student race (RACE) English as native language (BYS67)	Census region (BYREGION) Urbanicity (BYURBAN) Father’s educational attainment (FATHED) Mother’s educational attainment (MOTHED) Number persons dependent on parent (BYP06)
Mother’s educational attainment (MOTHED)	Student race (RACE) Parental aspirations for student postsecondary achievement (BYP79) Mother’s birthplace (BYP17)	Census region (BYREGION) Urbanicity (BYURBAN) Family composition (BYFCOMP) Student educational expectations (BYS56) Computer in home (BYS84C)
Mother’s occupation (OCCUMOTH)	Student race (RACE) Mother’s educational attainment (MOTHED) Mother’s birthplace (BYP17)	Census region (BYREGION) Urbanicity (BYURBAN) Family composition (BYFCOMP) Student educational expectations (BYS56) Parental aspirations for student postsecondary achievement (BYP79) Computer in home (BYS84C)
Father's educational attainment (FATHED)	Student race (RACE) Parental aspirations for student postsecondary achievement (BYP79) Father's birthplace (BYP20)	Census region (BYREGION) Urbanicity (BYURBAN) Family composition (BYFCOMP) Student educational expectations (BYS56) Computer in home (BYS84C)

See notes at end of table.

Table F–2. Imputation classes and sort variables for variables imputed by weighted sequential hot deck imputation—Continued

Imputation variable	Imputation class variables	Sort variables
Father's occupation (OCCUFATH)	Student race (RACE) Father's educational attainment (FATHED) Father's birthplace (BYP20)	Census region (BYREGION) Urbanicity (BYURBAN) Family composition (BYFCOMP) Student educational expectations (STEXPECT) Parental aspirations for student postsecondary achievement (PARASPIR) Computer in home (BYS84C)
Household income (INCOME)	Mother's educational attainment (MOTHED) Father's educational attainment (FATHED) Family composition (BYFCOMP)	Census region (BYREGION) Urbanicity (BYURBAN) Student race (RACE) Father's occupation (OCCUFATH) Number of earners contributing to family income (BYP86)

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table F–3. Variables included in multiple imputation model for student ability estimates for reading and mathematics

Imputation variable	Variables included in multiple imputation model
Student ability estimates (theta) for reading and mathematics	School type (BYSCTRL)
	Census region (BYREGION)
	Urbanicity (BYURBAN)
	Student sex (SEX)
	Student race (RACE)
	English as native language (BYS67)
	Mother's occupation (OCCUMOTH)
	Father's occupation (OCCUFATH)
	Student educational expectations (BYS56)
	Parental aspirations for student postsecondary achievement (BYP79)
	Mother's educational attainment (MOTHED)
	Father's educational attainment (FATHED)
	Household income (INCOME)
	Family composition (BYFCOMP)

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table F–4. Distribution of variables before and after imputation: 2002

Variable	Before imputation		After imputation		Significance (at 0.05)
	Sample size	Weighted percent	Sample size	Weighted percent	
Student sex (SEX) Total	**15,352**	100.0	**15,362**	100.0	
Male	7,640	50.5	7,646	50.5	*
Female	7,712	49.5	7,716	49.6	
Student race (RACE)	**15,355**	100.0	**15,362**	100.0	
American Indian or Alaska Native	131	0.9	131	0.9	
Asian	1,403	3.8	1,403	3.8	
Black or African American, non Hispanic	2,033	14.4	2,033	14.4	
Hispanic, no race specified	1,001	7.2	1,001	7.2	
Hispanic, race specified	1,234	8.8	1,233	8.8	
Multiracial, non-Hispanic	742	4.3	742	4.3	
Native Hawaiian or other Pacific Islander	62	0.2	62	0.2	
White, non-Hispanic	8,749	60.4	8,757	60.4	
Student Hispanic origin (HISPANIC)	**2,167**	100.0	**2,234**	100.0	
Mexican, Mexican-American, Chicano	1,421	69.3	1,466	69.4	
Cuban	77	3.2	80	3.2	
Dominican	67	3.8	68	3.7	
Puerto Rican	284	11.5	296	11.6	
Central American	152	6.3	155	6.2	
South American	166	6.0	169	6.0	
Student Asian origin (ASIAN)	**1,671**	100.0	**1,788**	100.0	
Chinese	351	20.1	375	20.2	
Filipino	263	20.5	284	20.0	
Japanese	128	8.5	137	8.4	
Korean	268	15.7	279	15.3	
Southeast Asian	411	20.6	443	21.2	
South Asian	250	14.7	270	14.9	
English as native language (BYS67)	**15,027**	100.0	**15,362**	100.0	*
English is native language	12,502	86.2	12,766	86.1	*
English not native language	2525	13.8	2596	14.0	*
Type of school program (BYS26)	**14,368**	100.0	**15,362**	100.0	
General	5,034	38.4	5,419	38.6	*
College preparatory-academic	7,920	50.9	8,439	50.7	*
Vocational-including technical/business	1,414	10.8	1504	10.8	*
Student educational expectations (BYS56)	**13,552**	100.0	**13,901**	100.0	*
Less than high school graduation	113	0.9	128	1.0	
High school graduation or GED only	930	7.8	999	8.1	*
Attend or complete a 2-year school	867	7.2	888	7.1	
Attend college, but not complete a 4-year degree	550	4.3	565	4.3	

See notes at end of table.

Table F–4. Distribution of variables before and after imputation: 2002—Continued

Variable	Before imputation: Sample size	Before imputation: Weighted percent	After imputation: Sample size	After imputation: Weighted percent	Significance (at 0.05)
Graduate from college	5,329	39.8	5,455	39.7	
Obtain a master's degree or equivalent	3,130	22.0	3,183	21.9	*
Obtain a PhD, MD, or other advanced degree	2,633	17.9	2,683	17.8	
Parental aspirations for student postsecondary achievement (BYP79)	**13,183**	100.0	**15,362**	100.0	*
Less than high school graduation	11	0.1	13	0.1	
High school graduation or GED only	438	3.8	543	4.0	*
Attend or complete a 2-year school	980	8.3	1,178	8.6	*
Attend college, but not complete a 4-year degree	125	100.0	145	100.0	*
Graduate from college	5,812	45.4	6,790	45.3	*
Obtain a master's degree or equivalent	2,773	20.1	3,200	19.9	*
Obtain a Ph.D., M.D., or other advanced degree	3,044	21.4	3,493	21.2	*
Family composition (BYFCOMP)	**13,487**	100.0	**15,362**	100.0	
Mother and father	8,111	57.3	9,131	56.8	*
Mother and male guardian	1,627	13.3	1,881	13.4	*
Father and female guardian	422	3.1	494	3.2	*
Two guardians	227	1.8	266	1.9	*
Mother only	2,376	19.0	2,755	19.1	*
Father only	400	3.1	454	3.2	*
Female guardian only	159	1.2	191	1.3	*
Male guardian only	40	0.2	48	0.2	
Lives with student less than half time	125	0.8	142	0.9	*
Mother's educational attainment (MOTHED)	**14,764**	100.0	**15,362**	100.0	*
Did not finish high school	1,821	12.9	1,933	13.2	*
Graduated from high school or GED	3,939	27.8	4,126	27.9	*
Attended 2-year school, no degree	1,783	13.1	1,856	13.1	*
Graduated from 2-year school	1,583	11.2	1,633	11.2	*
Attended college, no 4-year degree	1,556	10.5	1,595	10.3	*
Graduated from college	2,747	16.7	2,837	16.6	*
Completed master's degree or equivalent	1,034	6.0	1,066	6.0	*
Completed Ph.D., M.D., advanced degree	301	1.7	316	1.7	*
Mother's occupation (OCCUMOTH)	**14,514**	100.0	**15,362**	100.0	*
No job for pay	589	3.5	606	3.3	*
Clerical	2,348	16.7	2,480	16.7	*
Craftsperson	320	2.3	338	2.3	*
Farmer, farm manager	83	0.7	84	0.6	
Homemaker	616	4.3	761	5.0	*
Laborer	658	4.9	685	4.8	*
Manager, administrator	1,585	10.9	1,670	10.9	*
Military	28	0.2	29	0.2	

See notes at end of table.

Table F–4. Distribution of variables before and after imputation: 2002—Continued

Variable	Before imputation		After imputation		Significance (at 0.05)
	Sample size	Weighted percent	Sample size	Weighted percent	
Operative	608	4.5	638	4.4	*
Professional A	2,113	13.9	2,188	13.6	*
Professional B	589	3.8	606	3.7	*
Proprietor, owner	342	2.3	365	2.3	*
Protective service	106	0.8	114	0.7	
Sales	631	4.4	659	4.3	*
School teacher	984	6.5	1,009	6.3	*
Service	2,178	15.5	2,362	15.9	*
Technical	736	5.1	768	5.0	*
Father's educational attainment (FATHED)	**13,847**	100.0	**15,362**	100.0	*
Did not finish high school	1,792	13.6	2,040	13.9	*
Graduated from high school or GED	3,849	29.9	4,335	30.2	*
Attended 2-year school, no degree	1,298	9.9	1,450	9.9	*
Graduated from 2-year school	1,091	8.3	1,203	8.2	*
Attended college, no 4-year degree	1,294	9.4	1,426	9.3	*
Graduated from college	2,526	16.9	2,749	16.8	*
Completed master's degree or equivalent	1,187	7.5	1,289	7.4	*
Completed Ph.D., M.D., advanced degree	810	4.5	870	4.4	*
Father's occupation (OCCUFATH)	**13,147**	100.0	**15,362**	100.0	*
No job for pay	155	0.8	177	0.8	*
Clerical	314	2.5	365	2.5	*
Craftsperson	1,635	13.5	1,941	13.5	*
Farmer, farm manager	249	2.2	289	2.1	*
Homemaker	271	2.2	392	2.7	*
Laborer	1,327	10.7	1,615	11.1	*
Manager, administrator	2,007	15.0	2,264	14.5	*
Military	173	1.3	202	1.3	*
Operative	1,465	11.9	1,773	12.2	*
Professional A	1,449	10.3	1,636	10.0	*
Professional B	831	5.0	904	4.8	*
Proprietor, owner	817	5.9	930	5.8	*
Protective service	435	3.4	522	3.4	*
Sales	699	5.4	802	5.3	*
School teacher	200	1.5	216	1.4	*
Service	504	3.7	619	3.9	*
Technical	616	4.8	715	4.8	*
Household income (INCOME)	**11,907**	100.0	**15,362**	100.0	
None	56	0.4	73	0.4	*
$1000 or less	123	1.1	169	1.2	*
$1,001 – $5,000	214	1.8	285	1.8	*
$5,001 – $10,000	248	2.1	323	2.2	*

See notes at end of table.

Table F–4. Distribution of variables before and after imputation: 2002—Continued

Variable	Before Imputation		After Imputation		Significance (at 0.05)
	Sample size	Weighted percent	Sample size	Weighted percent	
$10,001 –- $15,000	498	4.3	659	4.4	*
$15,001 – $20,000	566	5.0	746	5.0	*
$20,001 – $25,000	694	6.2	940	6.5	*
$25,001 – $35,000	1,378	12.2	1,804	12.4	*
$35,001 – $50,000	2,203	19.3	2,882	19.6	*
$50,001 – $75,000	2,447	21.0	3,139	20.7	*
$75,001 – $100,000	1,641	13.2	2,064	12.9	*
$100,001 – $200,000	1,391	10.5	1,725	10.1	*
$200,001 or more	448	2.8	553	2.7	*

*A *t* test comparing the weighted percents before and after imputation showed a significant difference at the 0.05(c-1) level, where c is the number of categories within the primary variable.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

This page is intentionally blank.

Appendix G
Base Year Codebooks

This page is intentionally blank.

Appendix G
Base Year Codebooks

Web-published hardcopy codebooks are available as PDF files at: http://nces.ed.gov/surveys/els2002/

This page is intentionally blank.

Appendix H
Cross-Cohort Comparisons

This page is intentionally blank.

Appendix H
Cross-Cohort Comparisons

H.1 Cross-Cohort Comparison Crosswalks

ELS:2002 base year data can be used in cross-cohort (intercohort) comparisons with the sophomore cohorts of the High School and Beyond (HS&B) longitudinal study in 1980 and the National Education Longitudinal Study of 1988 (NELS:88) in 1990. This appendix contains four crosswalks designed to identify ELS:2002 variables that also appear on the HS&B 1980 sophomore cohort or NELS:88 1990 data sets. The four crosswalks encompass the student, parent, teacher, and school administrator components. Some items identified in the crosswalks are only approximate matches, and for these, analysts should judge whether they are sufficiently comparable for purposes of the analysis at hand. In other cases, question stems and response options correspond exactly across questionnaires. All 1980 and 2002 participants are by definition sophomores. However, for NELS:88 (1990), the subset of participants who were sophomores at the time must be invoked through use of the sophomore cohort flag.

While the three studies have been designed to produce comparable results (both to each other, and, at 12th grade, to the National Longitudinal Study of the High School Class of 1972 [NLS-72]), there are also differences between them that may affect the comparability of estimates. Analysts should be aware of and take account of these several factors. In particular, there are differences in sample eligibility and sampling rates, differences in response rates, and some differences in key classification variables, such as race/ethnicity.

Quite similar definitions were used in deciding issues of school eligibility across the studies. Differences in student sampling eligibility, however, are more problematic. Although the target population is essentially the same across the studies—all sophomores who can validly be assessed or at 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.

Not all students are able to meaningfully respond to research instruments such as the assessments and questionnaires administered in the three studies. Some students are too limited in their English language proficiency to do so, while others may be precluded from participation by a physical or mental disability. HS&B excluded as ineligible students with such barriers to participation, although an overall exclusion rate has not been documented. In NELS:88, 5.3 percent of the base year eighth-grade sample was excluded for such reasons (this figure is similar to the exclusion rate for eighth grade in the National Assessment of Educational Progress [NAEP] in similar subjects in the same period). However, a sample of the NELS:88 ineligible students was followed over time, and some students whose status changed were incorporated into the first follow-up, from which the NELS:88 sophomore cohort is drawn. In ELS:2002, no students were classified as ineligible as such, though some were exempted from completing the questionnaire or test, and others were tested under circumstances in which they were provided

with special accommodations. The overall rate of instrument-exempted sophomores in ELS:2002 is quite low, just over 1 percent. Contextual information was collected for these individuals. The instrument-exempted students are considered to be part of the study, but do not appear on the public-use file.

Differences in sampling rates and sample sizes across the studies will also affect power of generalization. Asian students, for example, have been oversampled in NELS:88 and in ELS:2002, but not in HS&B, where their numbers were quite small. Also, although Catholic schools were oversampled in all three studies, HS&B had few (only 38) private non-Catholic schools.

Response rates also differ somewhat across the studies, although nonresponse-adjusted weights were generated for each of the cohorts. At the school level, response rates were somewhat higher in HS&B and NELS:88 than in ELS:2002. School nonresponse bias analyses were performed for each study and may be found in the study documentation. At the student level, participation in ELS:2002 was higher than for HS&B sophomores, but lower than the rate for NELS:88. Of the HS&B sophomore cohort in 1980, 84 percent completed the student questionnaire and 77 percent completed the cognitive tests. For the NELS:88 sophomores in 1990, 94 percent completed the student questionnaire and 90 percent completed cognitive tests. In ELS:2002, 87 percent of sophomores completed a questionnaire and 83 percent also completed one or more assessments (of course coverage rates, the proportion of questionnaire completers with test data, are higher than overall response rates, e.g., 95 percent of NELS:88 student questionnaire completers have test data, and 95 percent of ELS:2002 questionnaire completers have assessment data as well).

In some cases, federal race definitions or preferences for the means by which ethnicity and race data are to be collected have changed. In HS&B and NELS:88, students were asked to mark one race only. Based on revised race-reporting guidelines issued by the Office of Management and Budget (OMB), a new race category has been added, and, more important, students are now allowed to mark all that apply, thus generating a further category, multiracial.

The new race category is Native Hawaiian or Other Pacific Islander. For purposes of cross-cohort comparisons, cases identified in ELS:2002 as “Native Hawaiian or Other Pacific Islander” should be combined with the category “Asian” to achieve comparability with HS&B and NELS:88.

However, for students who considered themselves to be multiracial and marked more than one race, there is no ready means to map them back into a one-race scheme. With 5 race categories, and values based on a single race reported, none reported, the 10 possible combinations of 2 races, 10 possible combinations of 3 races, the 5 possible combinations of 4 races, and the possibility of a combination of all 5 races, there are 32 separate race categories. When race is crossed by ethnicity (race by Hispanic or not Hispanic), there are 64 possible race-Hispanic ethnicity combinations. It is impossible to know, for example, whether a student who marked White and Black in ELS:2002 would have marked White, or have marked Black, in NELS:88, in which only one race was allowed. There are over 700 non-Hispanic multiracial sophomores recorded in the ELS:2002 base year data set, but the distorting effect on cross-cohort estimation is likely to be greatest for small population subgroups with many claimants to

multiple race, such as the American Indian category. It should also be noted that weights were created for racial groups on the basis of the school's classification of their sophomore enrollees (rosters normally assigned each student to a single race), not on the self-report from the student questionnaire. Analysts should be cautious, then, about conclusions concerning racial subgroup trends between the sophomores of 1980, 1990, and 2002.

Other key classification variables have been constructed *to the extent possible* in the same way in ELS:2002 as in the prior studies, although in many cases (in ELS:2002 only) there are imputed versions of the variable as well as the original version with the various types of missing data categorized by reserve code. The socioeconomic status (SES) variable offers a good example of how subtle differences may exist between the same variable in different studies, despite efforts to maximize cross-cohort consistency of measures. Although cross-cohort comparisons between ELS:2002 and NLS-72 and the senior cohort of HS&B will not be possible until the 2004 first follow-up, it may be useful to provide a more inclusive picture of the SES composite. Continuities and differences in SES constituents and construction in the three prior studies are summarized below in Table H-1. Table H-2 summarizes the elements comprising the SES measure in ELS:2002.

Table H–1. Socioeconomic composite, the National Education Longitudinal Study of 1988 as compared to the National Longitudinal Study of the High School Class of 1972 and the High School and Beyond longitudinal study

NLS-72, HS&B (student reported)	NELS:88 (parent reported)	NELS:88 student survey substitutions
Father's occupation	Father's occupation	Father's occupation
	Mother's occupation	Mother's occupation
Father's education	Father's education	Father's education
Mother's education	Mother's education	Mother's education
Family income	Family income	Household items
Household items	—	

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table H–2. Socioeconomic composite, the Education Longitudinal Study of 2002

Preferred source (parent reported)	Student report substitution if missing from parent	Imputed if still missing
Father's occupation	Father's occupation	Father's occupation
Mother's occupation	Mother's occupation	Mother's occupation
Father's education	Father's education	Father's education
Mother's education	Mother's education	Mother's education
Family income	—	Family income

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

ELS:2002 largely follows the NELS:88 model above in that in both studies the composite is based on five equally weighted, standardized components: father's education, mother's education, family income, father's occupation, and mother's occupation. Parent data are used to construct this variable. Student data are substituted where parent data are missing. However, for

parent education and occupation, where both parent and student reports are missing, ELS:2002 education and occupation values are imputed. Family income was not asked of students. While in NELS:88 a student-provided household item index, which served as an income proxy, was substituted when income data were missing, a different procedure was followed in ELS:2002. When parent data on income were missing, income was statistically imputed.

Some differences across the studies are based on differences in design. The studies had different starting points. NLS-72 student respondents were high school seniors, HS&B base year respondents were sophomores or seniors, and NELS:88 base year respondents were eighth graders. ELS:2002 student respondents were sophomores. A parent interview was sought for all NELS:88 and ELS:2002 base year student respondents. HS&B had a parent survey, but it only encompassed a subsample of student respondents. NLS-72 had no parent survey at all. Since the quality of reporting on parental occupation and education increases with student age or grade, it may be of concern whether reports were gathered at grade 8, 10, or 12. However, since parent reports are markedly superior to student reports in these matters, it may be of concern that only in NELS:88 and ELS:2002 are the data primarily parent reported.

Some differences reflect changing social circumstances over time. For example, many fewer mothers worked in 1972. The importance of gathering information about maternal occupation increased with the passage of time and the increasing labor market participation of American females. The household items list has been revised for each survey. For NLS-72, owning a color television discriminated between people of various income levels. By the time of HS&B, 8 years later, this was no longer so. By 2002, HS&B items such as ownership of a typewriter had ceased to function as good proxies for family income, while other items, such as access to the Internet or having a digital video disc player did.[1] While items differ across the index over time, in each case the items are those that are needed to provide a measure that has a reasonable correlation with income. Another area where change over time is possible is in occupations and their relative prestige. To accommodate this factor, two sets of prestige scores were drawn upon in NELS:88: the 1961 Duncan socioeconomic indicator measure that had been employed in NLS-72 and HS&B, as well as a 1989 revision by Nakao and Treas (1992). The same strategy has been employed in ELS:2002.

One difference between the SES variable in ELS:2002 and in prior studies arises from the use of imputation in ELS:2002. Since 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) just under 10 percent. The availability of imputed variables (including both key classification variables and achievement test scores) also poses a novel question for analysts interested in intercohort comparisons. Since imputed values are flagged, it is the analyst's choice whether or not to employ them. If the imputed variables are used, they should have the effect of improving cross-sectional estimation. On the other hand, since imputation was not used in the prior studies, it is also possible that use of ELS:2002 imputed values might decrease comparability of results across studies. To explore the issue of the magnitude of the effect of imputation on comparative bivariate and multivariate analysis, the forthcoming NCES trend

[1] The household items were asked in ELS:2002, but the index was not used in the creation of SES, since missing income data were imputed.

report on sophomores will compare imputed and unimputed ELS:2002 estimates, including estimates based on an SES composite using the household items index substitution, and an SES composite based on parent data with missings imputed.

The crosswalk below links ELS:2002 base year student questionnaire items with similar items from two previous NCES sophomore high school cohort questionnaires: the NELS:88 first follow-up questionnaire (1990) and the HS&B base year sophomore questionnaire (1980). This crosswalk will facilitate analyses of trends among sophomore high school students spanning 22 years. Linked questions may be identical in content and format or may differ in one or more ways: the question, item, or response wording; the order in which response options were presented; the manner in which the data were collected (e.g., categorical response option versus open-ended response fields, instructions to mark one versus mark all that apply); and the population to which the question applies may have changed. Therefore, it is strongly recommended that analysts review documentation to determine if linked questions are appropriate for their purpose.

This is best illustrated by way of example. Question 52 on the ELS:2002 base year questionnaire reads: "Do you plan to continue your education right after high school or at some time in the future? Question 51 on the NELS:88 first follow-up questionnaire is similar, but refers to college rather than education more generally ("Do you plan to go to college after you graduate from high school?"). The parallel question on the HS&B base year questionnaire (112) pertains to college, but does not presume high school graduation ("Do you plan to go to college at some time in the future?"). Whether these questions are comparable depends on the research question and the analyst's objective. To take another example, the item "I don't feel safe at this school" is common to all three studies. However, in 1980, it was asked dichotomously. In 1990 and 2002, it was asked with four response options: strongly agree, agree, disagree, and strongly disagree. The NELS:88 and ELS:2002 responses can be directly compared. The time line can be extended back to HS&B if, and only if, the analyst is willing to collapse the NELS:88 and ELS:2002 "strongly agree" and "agree" categories, and merge "disagree" with "strongly disagree."

Table H–3. Intercohort student questionnaire crosswalk

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
14	N2♦	84	Sex of Student
15	N8A♦	91	Student is Hispanic
16	N8C♦	91	Student's Hispanic subgroup
17A	N8A♦	90	Student is White
17B	N8A♦	90	Student is Black/African American
17C	N8A♦	90	Student is Asian
17D	N8A♦	90	Student is Native Hawaiian/Pacific Islander
17E	N8A♦	90	Student is American Indian/Alaska Native
18	N8A♦	91	Student's Asian subgroup
20A	7A	-	Students get along well with teachers
20B	7B	-	There is real school spirit
20C	7E	-	Students friendly with other racial groups
20D	7F	-	Other students often disrupt class
20E	7G	-	The teaching is good
20F	7H	-	Teachers are interested in students
20G	7I	-	Teachers praise effort
20H	7J	-	In class often feels put down by teachers
20I	7K	-	In class often feels put down by students
20J	7M	66F	Does not feel safe at this school
20K	7N	-	Disruptions get in way of learning
20L	7O	-	Misbehaving students often get away with it
20M	-	-	There are gangs in school
20N	-	-	Racial/ethnic groups often fight
21A	-	-	Everyone knows what school rules are
21B	-	-	School rules are fair
21C	-	-	Punishment same no matter who you are
21D	-	-	School rules are strictly enforced
21E	-	-	Students know punishment for broken rules
22A	9A	-	Had something stolen at school
22B	9B	-	Someone offered drugs at school
22C	9C	-	Someone threatened to hurt 10th grader at school
22D	9D	-	Got into a physical fight at school
22E	-	-	Someone hit 10th grader
22F	-	-	Someone forced money/things from 10th grader
22G	-	-	Someone damaged belongings
22H	-	-	Someone bullied or picked on 10th grader
23A	8C	-	Won an academic honor
23B	8E	-	Recognized for good attendance
23C	8F	-	Recognized for good grades
23D	8I	-	Received community service award
23E	8D	-	Participated in science/math fair
23F	8J	-	Participated in voc/tech skills competition
24A	10A	18	How many times late for school
24B	10B	-	How many times cut/skip classes
24C	-	17	How many times absent from school
24D	10C	-	How many times got in trouble
24E	10D	-	How many times put on in-school suspension

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
24F	10E	-	How many times suspended/put on probation
24G	10F	-	How many times transferred for disciplinary reasons
25AA	-	-	1st friend's sex
25BA	-	-	1st friend is Hispanic
25CAA	-	-	1st friend is White
25CAB	-	-	1st friend is Black/African American
25CAC	-	-	1st friend is Asian
25CAD	-	-	1st friend is Native Hawaiian/Pacific Islander
25CAE	-	-	1st friend is American Indian/Alaska Native
25DA	-	-	1st friend's grade level at school
25EA	-	-	Importance of grades to 1st friend
25FA	-	-	10th grader knows 1st friend's parents
25GA	-	-	Parents know 1st friend's parents
25AB	-	-	2nd friend's sex
25BB	-	-	2nd friend is Hispanic
25CBA	-	-	2nd friend is White
25CBB	-	-	2nd friend is Black/African American
25CBC	-	-	2nd friend is Asian
25CBD	-	-	2nd friend is Native Hawaiian/Pacific Islander
25CBE	-	-	2nd friend is American Indian/Alaska Native
25DB	-	-	2nd friend's grade level at school
25EB	-	-	Importance of grades to 2nd friend
25FB	-	-	10th grader knows 2nd friend's parents
25GB	-	-	Parents know 2nd friend's parents
25AC	-	-	3rd friend's sex
25BC	-	-	3rd friend is Hispanic
25CCA	-	-	3rd friend is White
25CCB	-	-	3rd friend is Black/African American
25CCC	-	-	3rd friend is Asian
25CCD	-	-	3rd friend is Native Hawaiian/Pacific Islander
25CCE	-	-	3rd friend is American Indian/Alaska Native
25DC	-	-	3rd friend's grade level at school
25EC	-	-	Importance of grades to 3rd friend
25FC	-	-	10th grader knows 3rd friend's parents
25GC	-	-	Parents know 3rd friend's parents
26	20	1	High school program–student self-report
27A	66A	-	Classes are interesting and challenging
27B	66B	-	Satisfied by doing what expected in class
27C	66C	-	Has nothing better to do than school
27D	66D	-	Education is important to get a job later
27E	66E	-	School is a place to meet friends
27F	66F	-	Plays on a team or belongs to a club
27G	-	-	Learns skills for job in school
27H	66G	-	Teachers expect success in school
27I	-	-	Parents expect success in school
28	-	-	How much likes school
29A	32A	-	How often reviews work in math class
29B	-	-	How often listens to math teacher lecture

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
29C	32C	-	How often copies math teacher's notes from board
29D	32B	-	How often uses books besides math textbooks
29E	32D	-	How often does problem-solving in math class
29F	32G	-	How often uses calculators in math class
29G	-	-	How often uses graphing calculators in math class
29H	32E	-	How often uses computers in math class
29I	32I	-	How often explains work to math class orally
29J	32H	-	How often participates in student math discussions
30	-	-	Uses computers in math class
31A	-	-	How often uses computers to review math work
31B	-	-	How often uses computers to solve math problems
31C	-	-	How often uses computers for graphing in math class
31D	-	-	How often uses computers to practice math drills
31E	-	-	How often uses computers to analyze data in math class
31F	-	-	How often uses computers to apply learning in math class
31G	-	-	How often math teacher uses computer to instruct one-on-one
31H	-	-	How often math teacher uses computer to show new topics
32AA	-	-	Used computer in 9th-grade fall English
32BA	-	-	Used computer in 9th-grade spring English
32CA	-	-	Used computer in 9th-grade fall science
32DA	-	-	Used computer in 9th-grade spring science
32EA	-	-	Used computer in 9th-grade fall math
32FA	-	-	Used computer in 9th-grade spring math
32GA	-	-	Used computer in 9th-grade fall social studies
32HA	-	-	Used computer in 9th-grade spring social studies
32AB	-	-	Uses computer in 10th-grade fall English
32BB	-	-	Uses computer in 10th-grade spring English
32CB	-	-	Used computer in 10th-grade fall science
32DB	-	-	Uses computer in 10th-grade spring science
32EB	-	-	Used computer in 10th-grade fall math
32FB	-	-	Uses computer in 10th-grade spring math
32GB	-	-	Used computer in 10th-grade fall social studies
32HB	-	-	Uses computer in 10th-grade spring social studies
33A	34E	-	Ever in Advanced Placement program
33B	-	-	Ever in International Baccalaureate program
33C	-	-	Ever in part-time program at regional vocational school
33D	34A	13A	Ever in a remedial English class
33E	34B	13B	Ever in a remedial math class
33F	34C	13E	Ever in bilingual/bicultural class
33G	34D	-	Ever in English as Second Language program
33H	34H	-	Ever in dropout prevention program
33I	34F, G	13H, I	Ever in special education program
33J	-	-	Ever in distance learning course
33K	-	-	Ever in career academy
33L	-	-	Ever in program to help prepare for college
34A	36A	15	Hours/week spent on homework in school
34B	36A	15	Hours/week spent on homework out of school

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
35A	36B	-	Hours/week spent on math homework in school
35B	36B	-	Hours/week spent on math homework out of school
36A	36D	-	Hours/week spent on English homework in school
36B	36D	-	Hours/week spent on English homework out of school
37	38	-	Importance of good grades to student
38A	40A	16A	How often goes to class without pencil/paper
38B	40B	16B	How often goes to class without books
38C	40C	16C	How often goes to class without homework done
39A	41AA	34A	Played intramural baseball
39B	41AA	34A	Played intramural softball
39C	41AB	34A	Played intramural basketball
39D	41AC	34A	Played intramural football
39E	41AD	34A	Played intramural soccer
39F	41AF	34A	Played other intramural team sport
39G	41AG	34A	Played an individual intramural sport
39H	41AH, AI	34B	On intramural cheerleading/drill team
40AA	41AA	34A	No interscholastic baseball
40AB	41AA	34A	Did not participate in interscholastic baseball
40AC	41AA	34A	Played junior varsity baseball
40AD	41AA	34A	Played varsity baseball
40AE	41AA	34A	Varsity baseball captain/co-captain
40BA	41AA	34A	No interscholastic softball
40BB	41AA	34A	Did not participate in interscholastic softball
40BC	41AA	34A	Played junior varsity softball
40BD	41AA	34A	Played varsity softball
40BE	41AA	34A	Varsity softball captain/co-captain
40CA	41AB	34A	No interscholastic basketball
40CB	41AB	34A	Did not participate in interscholastic basketball
40CC	41AB	34A	Played junior varsity basketball
40CD	41AB	34A	Played varsity basketball
40CE	41AB	34A	Varsity basketball captain/co-captain
40DA	41AC	34A	No interscholastic football
40DB	41AC	34A	Did not participate in interscholastic football
40DC	41AC	34A	Played junior varsity football
40DD	41AC	34A	Played varsity football
40DE	41AC	34A	Varsity football captain/co-captain
40EA	41AD	34A	No interscholastic soccer
40EB	41AD	34A	Did not participate in interscholastic soccer
40EC	41AD	34A	Played junior varsity soccer
40ED	41AD	34A	Played varsity soccer
40EE	41AD	34A	Varsity soccer captain/co-captain
40FA	41AF	34A	No other interscholastic team sport
40FB	41AF	34A	Did not participate in other interscholastic team sport
40FC	41AF	34A	Played on other junior varsity team
40FD	41AF	34A	Played on other varsity team
40FE	41AF	34A	Varsity captain/co-captain for other team sport
40GA	41AG	34A	No interscholastic individual sport
40GB	41AG	34A	Did not participate in interscholastic individual sport

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
40GC	41AG	34A	Played junior varsity individual sport
40GD	41AG	34A	Played varsity individual sport
40GE	41AG	34A	Varsity captain/co-captain for individual sport
40HA	41AH, AI	34B	No interscholastic cheerleading/drill team
40HB	41AH, AI	34B	Did not participate on interscholastic cheerleading/drill team
40HC	41AH, AI	34B	Participated on junior varsity cheerleading/drill team
40HD	41AH, AI	34B	Participated on varsity cheerleading/drill team
40HE	41AH, AI	34B	Varsity cheerleading/drill team captain/co-captain
41A	41BA	34D, E	Participated in school band or chorus
41B	41BB	-	Participated in school play or musical
41C	41BC	-	Participated in student government
41D	41BD	-	Participated in academic honor society
41E	41BE	-	Participated in school yearbook or newspaper
41F	41BF	-	Participated in school service clubs
41G	41BG	34G	Participated in school academic clubs
41H	41BH	34F	Participated in school hobby clubs
41I	41BI	34H	Participated in school vocational clubs
42	42	-	Hours/week spent on extracurricular activities
43	43	-	Hours/week spent reading outside of school
44A	44A	47A	How often visits with friends at local hangout
44B	44C	-	How often works on hobbies
44C	44H	-	How often performs community services
44D	44I	47D	How often drives or rides around
44E	44J	47E	How often talks on phone with friends
44F	44M	-	How often takes music, art, language class
44G	44N	-	How often takes sports lessons
44H	44F	-	How often plays non-school sports
45A	-	-	How often uses computer for fun
45B	-	-	How often uses computer for school work
45C	-	-	How often uses computer to learn on own
46A	-	-	Hours/day on computer for school work
46B	-	-	Hours/day on computer other than for school
47A	-	-	How often uses computer at home
47B	-	-	How often uses computer at school
47C	-	-	How often uses computer at public library
47D	-	-	How often uses computer at friend's house
47E	-	-	How often uses computer at another place
48A	45A	48	Hours/day spent watching TV/DVD on weekdays
48B	45B	-	Hours/day spent watching TV/DVD on weekends
49A	-	-	Hours/day plays video/computer games on weekdays
49B	-	-	Hours/day plays video/computer games on weekends
50	-	-	School has library media/resource center
51A	-	-	Use of school library for assignments
51B	-	-	Use of school library for in-school projects
51C	-	-	Use of school library for homework
51D	-	-	Use of school library for research papers
51E	-	-	Use of school library for leisure reading

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
51F	-	-	Use of school library to read magazines/newspapers
51G	-	-	Use of school library to read books for fun
51H	-	-	Use of school library for interests outside of school
51I	-	-	Use of school library for Internet access
52	-	-	How useful are school library reference materials
53A	-	-	How helpful is library staff with finding research resources
53B	-	-	How helpful is library staff with using databases
53C	-	-	How helpful is library staff with using Internet
54A	46A	61A	Importance of being successful in line work
54B	46B	61B	Importance of marrying right person/having happy family
54C	46C	61C	Importance of having lots of money
54D	46D	61D	Importance of having strong friendships
54E	46E	61E	Importance of being able to find steady work
54F	46F	-	Importance of helping others in community
54G	46G	61G	Importance of giving children better opportunities
54H	46H	61H	Importance of living close to parents/relatives
54I	46I	61I	Importance of getting away from this area
54J	46J	61J	Importance of working to correct inequalities
54K	46K	61K	Importance of having children
54L	46L	61L	Importance of having leisure time
54N	-	-	Importance of being expert in field of work
54O	-	-	Importance of getting good education
55A	50A, F	-	Plans to take the PSAT or PACT
55B	50B, C	-	Plans to take SAT or ACT
55C	50D	-	Plans to take Advanced Placement test
55D	50E	-	Plans to take the ASVAB
56	49	69	How far in school student thinks will get
57	51	112	Plans to continue education after high school
58	-	-	Type of school plans to attend
59A	-	-	Has gone to counselor for college entrance information
59B	-	-	Has gone to teacher for college entrance information
59C	-	-	Has gone to coach for college entrance information
59D	-	-	Has gone to parent for college entrance information
59E	-	-	Has gone to friend for college entrance information
59F	-	-	Has gone to sibling for college entrance information
59G	-	-	Has gone to other relative for college entrance information
59H	-	-	Has gone to college publications/websites for entrance information
59I	-	-	Has gone to college representatives for entrance information
59J	-	-	Has gone to college search guides for entrance information
59K	-	-	Did not go to any of these sources for college entrance information
60	-	-	Would like to play athletics in college
61	-	-	Hopes to receive athletic scholarship for college
62A	-	-	Does not like school
62B	-	-	Grades are not good enough
62C	-	-	Will not need more school for job
62D	-	-	Cannot afford school
62E	-	-	Would rather work and earn money
62F	-	-	Plans to be full-time homemaker

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
62G	-	-	Does not feel school is important
62H	-	-	Needs to support family
63	53A	-	Occupation expects to have after high school–verbatim
64	53B	68	Occupation expects to have at age 30–verbatim
65A	48B	70	How far in school mother wants 10th grader to go
65B	48A	-	How far in school father wants 10th grader to go
66A	47B	50B	Mother's desire for 10th grader after high school
66B	47A	50A	Father's desire for 10th grader after high school
66C	47C	50E	Friend's desire for 10th grader after high school
66D	47D	50E	Close relative's desire for 10th grader after high school
66E	47E	50C	School counselor's desire for 10th grader after high school
66F	47F	50D	Favorite teacher's desire for 10th grader after high school
66G	47G	-	Coach's desire for 10th grader after high school
67	N12*	11	English is student's native language
68	N12*	11	Student's native language
69A	-	18A, B	How often 10th grader speaks native language with mother
69B	-	18C, D	How often 10th grader speaks native language with father
69C	-	-	How often 10th grader speaks native language with siblings
69D	-	18G	How often 10th grader speaks native language with friends
70A	57A	19A	How well 10th grader understands spoken English
70B	57B	19B	How well 10th grader speaks English
70C	57C	19C	How well 10th grader reads English
70D	57D	19D	How well 10th grader writes English
71A	-	-	Participated in cooperative education
71B	-	-	Participated in internship
71C	-	-	Participated in job shadowing/worksite visits
71D	-	-	Participated in mentoring
71E	-	-	Participated in community service
71F	-	-	Participated in school-based enterprise
71G	-	-	Did not participate in these work-based learning experiences
72	84	-	Ever worked for pay not around house
73	-	24	Date last worked for pay
74	-	-	Date started current/most recent job
75	85	25	How many hours usually works a week
76	86	-	How many hours works on the weekend
77	87	27	Type of work does on current/most recent job
79	-		How got current/most recent job
80	-	-	How closely related job is to desired job after education
81A	N5B*	41	Mother/female guardian's occupation–verbatim
81B	-	41	Mother/female guardian's main job duties–verbatim
82A	N7B*	38	Father/male guardian's occupation–verbatim
82B	-	38	Father/male guardian's main job duties-verbatim
83A	N20B*	41	Mother's highest level of education
83B	N20A*	39	Father's highest level of education
84A	-	103B	Family has a daily newspaper
84B	-	-	Family has regularly received magazine
84C	-	-	Family has a computer

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
84D	-	-	Family has access to the Internet
84E	-	-	Family has DVD player
84F	-	103E	Family has an electric dishwasher
84G	-	-	Family has a clothes dryer
84H	-	103G	Family has more than 50 books
84I	-	103H	Has own room
84J	-	-	Family has fax machine
85A	100A	-	How often parents checks homework
85B	100B	-	How often parents help with homework
85C	100C	-	Special privileges given for good grades
85D	100D	-	Parents limit privileges due to poor grades
85E	100E	-	Required to work around the house
85F	100F	-	Parents limit TV watching or video games
85G	100G	-	Parents limit time with friends
86A	105A	-	How often discussed school courses with parents
86B	105B	-	How often discussed school activities with parents
86C	105C	-	How often discuss things studied in class with parents
86D	105D	-	How often discussed grades with parents
86E	105E	-	How often discussed transferring with parents
86F	105F	-	How often discussed prep for ACT/SAT with parents
86G	105G	-	How often discussed going to college with parents
86H	-	-	How often discussed current events with parents
86I	-	-	How often discussed troubling things with parents
87A	-	-	Gets totally absorbed in mathematics
87B	-	-	Thinks reading is fun
87C	-	-	Thinks math is fun
87D	-	-	Reads in spare time
87E	-	-	Gets totally absorbed in reading
87F	-	-	Mathematics is important
88A	-	-	Most people can learn to be good at math
88B	-	-	Have to be born with ability to be good at math
89A	-	-	Can do excellent job on math tests
89B	-	-	Can understand difficult math texts
89C	-	-	Can understand difficult English texts
89D	-	-	Studies to get a good grade
89E	-	-	Can learn something really hard
89F	-	-	Can understand difficult English class
89G	-	-	Remembers most important things when studies
89H	-	-	Studies to increase job opportunities
89I	-	-	Can do excellent job on English assignments
89J	-	-	Works as hard as possible when studies
89K	-	-	Can do excellent job on English tests
89L	-	-	Can understand difficult math class
89M	-	-	Can master skills in English class
89N	-	-	Can get no bad grades if decides to
89O	-	-	Keeps studying even if material is difficult
89P	-	-	Studies to ensure financially security

See notes at end of table.

Table H–3. Intercohort student questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 1st follow-up	HS&B: 1980 base year	ELS:2002 base year variable label
89Q	-	-	Can get no problems wrong if decides to
89R	-	-	Can do excellent job on math assignments
89S	-	-	Does best to learn what studies
89T	-	-	Can learn something well if wants to
89U	-	-	Can master math class skills
89V	-	-	Puts forth best effort when studying
90A	70A	-	Important to friends to attend classes regularly
90B	70B	-	Important to friends to study
90C	70C	-	Important to friends to play sports
90D	70D	-	Important to friends to get good grades
90E	70E	-	Important to friends to be popular with students
90F	70F	-	Important to friends to finish high school
90G	70G	-	Important to friends to have steady boy/girlfriend
90H	70I	-	Important to friends to continue education past high school
90J	70K	-	Important to friends to do community work
90K	70L	-	Important to friends to have job
90L	-	-	Important to friends to get together with friends
90M	70H	-	Important to friends to go to parties
90Q	-	-	Important to friends to make money
91	69	-	Number of close friends who dropped out
92A	-	-	Girls should have same opportunities in sports
92B	-	-	Some sports should be just for boys
92C	-	-	Girls should have own sports teams
92D	-	-	Girls should be on same sports teams as boys
94	68	-	Has close friends who were friends in 8th grade
96	-	-	Observed students betting on sports
97A	-	-	Bets were placed with friends
97B	-	-	Bets were placed with family members
97C	-	-	Bets were placed with bookie
97D	-	-	Bets were placed with a website
97E	-	-	Bets were placed through other means

♦ NELS:88 first follow-up question numbers preceded by "N" refer to questions on the New Participant Supplement, a brief questionnaire completed by first follow-up participants who did not participate in the base year (or who did not answer these key questions on the base year questionnaire). Corresponding data collected from base year participants were stored in the first follow-up New Participant variables.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table H–4. Intercohort parent questionnaire crosswalk

ELS:2002 base year	NELS:88 base year	NELS:88 2nd follow-up	ELS:2002 base year variable label
1	1A	1	Relationship to 10th grader
2	-	-	Biological/adoptive parent lives with 10th grader
3	-	-	Lives with a spouse or partner
4	1A	1	Spouse/partner's relationship to 10th grader
5	1B	2	How often 10th grader lives with respondent
6	2	6	Number of dependents
7A	-	9A	Number full/adoptive brothers live with 10th grader
7B	-	9A	Number half-brothers live with 10th grader
7C	-	9A	Number step-brothers live with 10th grader
7D	-	9B	Number full/adoptive sisters live with 10th grader
7E	-	9B	Number half-sisters live with 10th grader
7F	-	9B	Number step-sisters live with 10th grader
7G	-	9C	Number of 10th grader's children live with 10th grader
7H	-	9D	Number grandparents live with 10th grader
7I	-	9E	Number other relatives under 18 live with 10th grader
7J	-	9F	Number other relatives 18 or older live with 10th grader
7K	-	9G	Number non-relatives under 18 live with 10th grader
7L	-	9H	Number non-relatives 18 or older live with 10th grader
8	3A	102	Number of siblings 10th grader has
9	6	105	Number of siblings who dropped out of high school
10	7	7	Current marital status of parent respondent
11	8	106	Parent respondent's year of birth
12	9	107	Spouse/partner's year of birth
13	10	19	Parent is Hispanic
14	10B	21	Parent's Hispanic subgroup
15A	10	19	Parent is White
15B	10	19	Parent is Black or African American
15C	10	19	Parent is Asian
15D	10	19	Parent is Native Hawaiian/Pacific Islander
15E	10	19	Parent is American Indian/Alaska Native
16	10A	20	Parent's Asian subgroup
17	11	-	Whether 10th grader's mother's birthplace in US or elsewhere
18	12	-	Number of years ago mother came to US
19A	13	-	Mother's occupation before coming to US
19B	-	-	Mother's main job duties outside US
20	14	-	Whether 10th grader's father's birthplace in US or elsewhere
21	15	-	Number of years ago father came to US
22A	16	-	Father's occupation before coming to US
22B	-	-	Father's job main duties outside US
23	17	-	Whether 10th grader's birthplace in US or elsewhere
24	18	-	Number of years ago 10th grader came to US
25	19	-	10th grader attended school outside US
26A	20	-	10th grader completed kindergarten outside US

See notes at end of table.

Table H–4. Intercohort parent questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 2nd follow-up	ELS:2002 base year variable label
26B	20	-	10th grader completed 1st grade outside US
26C	20	-	10th grader completed 2nd grade outside US
26D	20	-	10th grader completed 3rd grade outside US
26E	20	-	10th grader completed 4th grade outside US
26F	20	-	10th grader completed 5th grade outside US
26G	20	-	10th grader completed 6th grade outside US
26H	20	-	10th grader completed 7th grade outside US
26I	20	-	10th grader completed 8th grade outside US
26J	-	-	10th grader completed 9th grade outside US
26K	-	-	10th grader completed 10th grade outside US
26L	20	-	10th grader did not complete any grades outside US
27	21	-	Grade student placed in when started school in US
28	-	22	English is parent respondent's native language
29	-	23	Native language of parent respondent
30A	-	24A	How often parent speaks native language with spouse/partner
30B	-	24B	How often parent speaks native language with children
30C	-	24C	How often parent speaks native language with other relatives
30D	-	24D	How often parent speaks native language with friends
31A	26A	25A	How well parent understands spoken English
31B	26B	25B	How well parent speaks English
31C	26C	25C	How well parent reads English
31D	26D	25D	How well parent writes English
32A	-	26A	Problems reading English books/magazines
32B	-	26B	Problems parent has filling out forms in English
32C	-	26C	Problems parent has understanding 10th grader's teachers
32D	-	26D	Problems parent has making self understood by teachers
32E	-	26E	Problems helping 10th grader with homework in English
33	29	-	Religious background of parent respondent
34A	30	-	Parent's highest level of education completed
34B	31	-	Spouse/partner's highest level of education completed
35A	-	-	Parent's mother's highest level of education
35B	-	-	Parent's father's highest level of education
35C	-	-	Spouse/partner's mother's highest level education
35D	-	-	Spouse/partner's father's highest level education
36	32	-	Parent working for pay during past week
37	32, 33A	-	Parent's current work status
38	33B	12	Whether parent ever held regular job for pay in US
39A	-	-	Parent's current/most recent job for pay in US
39B	-	-	Parent's main job duties
39C	34B	13	Parent's job description category
40	35	-	Spouse/partner working for pay during past week

See notes at end of table.

Table H–4. Intercohort parent questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 2nd follow-up	ELS:2002 base year variable label
41	35, 36A	-	Spouse/partner's current work status
42	36B	15	Whether spouse/partner ever held regular job for pay in US
43A	-	-	Spouse/partner's current/most recent job for pay in US
43B	-	-	Spouse/partner's main job duties
43C	37B	16	Spouse/partner's job description category
44A	38A	-	10th grader attended day care program
44B	38B	-	10th grader attended nursery/pre-school
44C	38C	-	10th grader attended Head Start program
44D	38D	-	10th grader attended kindergarten
45	40	33	Number times 10th grader changed schools other than promotions
46	44	-	10th grader ever held back a grade
47A	45A	-	10th grader held back because of parental request
47B	45B	-	10th grader held back because of school request
47C	45C	-	10th grader held back for other reason
48A	46	-	10th grader repeated kindergarten
48B	46	-	10th grader repeated 1st grade
48C	46	-	10th grader repeated 2nd grade
48D	46	-	10th grader repeated 3rd grade
48E	46	-	10th grader repeated 4th grade
48F	46	-	10th grader repeated 5th grade
48G	46	-	10th grader repeated 6th grade
48H	46	-	10th grader repeated 7th grade
48I	46	-	10th grader repeated 8th grade
48J	-	-	10th grader repeated 9th grade
48K	-	-	10th grader repeated 10th grade
49	-	-	Thinks 10th grader has disability
50A	47G	-	10th grader has specific learning disabilities
50B	47D	-	10th grader has speech/language impairments
50C	47I	-	10th grader has mental retardation
50D	47H	-	10th grader has emotional disturbance
50E	47B	-	10th grader has hearing impairments
50F	47E	-	10th grader has orthopedic impairments
50G	47A	-	10th grader has visual impairments
50H	47J	-	10th grader has other disability
51	50	35A	10th grader ever had behavior problem at school
52A	57A	43A	School contacted parent about poor performance
52B	57B	43B	School contacted parent about school program for year
52C	-	43C	School contacted parent about plans after high school
52D	57C	43D	School contacted parent about course selection
52E	-	43E	School contacted parent about poor attendance
52F	57E	43F	School contacted parent about problem behavior
52G	57E	43F	School contacted parent about positive/good behavior
52H	57F, H	43G	School contacted parent about fundraising/volunteer work

See notes at end of table.

Table H–4. Intercohort parent questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 2nd follow-up	ELS:2002 base year variable label
52I	-	43H	School contacted parent about helping with homework
52J	57G	-	School contacted parent to obtain information for records
53A	58A	44A	Parent contacted school about poor performance
53B	58B	44B	Parent contacted school about school program for year
53C	-	44C	Parent contacted school about plans after high school
53D	-	44D	Parent contacted school about course selection
53E	-	53E	Parent contacted school about poor attendance
53F	58C	53F	Parent contacted school about problem behavior
53G	58C	53F	Parent contacted school about positive/good behavior
53H	58D, F	53G	Parent contacted school about fundraising/volunteer work
53I	-	53H	Parent contacted school about helping with homework
53J	58E	-	Parent contacted school to provide information for records
54A	59A	-	Belong to parent-teacher organization
54B	59B	-	Attend parent-teacher organization meetings
54C	59C	-	Take part in parent-teach organization activities
54D	59D	-	Act as a volunteer at the school
54E	59E	-	Belong to other organization with parents from school
55A	-	-	How often check that homework completed
55B	-	-	How often discuss report card
55C	-	-	How often know whereabouts
55D	-	-	How often make/enforce school night curfews
56A	-	49A	Provide advice about selecting courses or programs
56B	-	49E	Provide advice about plans for college entrance exams
56C	-	49F	Provide advice about applying to college/school after hs
56D	-	49G	Provide advice about jobs to apply for after high school
56E	-	49H	Provide information about community/national/world events
56F	-	49I	Provide advice about things troubling 10th grader
57A	-	50A	Attended school activities with 10th grader
57B	-	50B	Worked on homework/school projects with 10th grader
57C	-	50C	Attended concerts/plays/movies with 10th grader
57D	-	50D	Attended sports events outside school with 10th grader
57E	-	50E	Attended religious services with 10th grader
57F	-	50F	Attended family social functions with 10th grader
57G	-	50G	Took day trips/vacations with 10th grader
57H	-	50H	Worked on hobby/played sports with 10th grader
57I	-	50I	Went shopping with 10th grader
57J	-	50J	Went to restaurants with 10th grader
57K	-	50K	Spent time talking with 10th grader
57L	-	50L	Did something else fun with 10th grader
58A	-	-	Most people can learn to be good at math-parent's opinion
58B	-	-	Must be born w/ability to be good at math-parent's opinion
59BA	62A	54A	1st friend attends same school

See notes at end of table.

Table H–4. Intercohort parent questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 2nd follow-up	ELS:2002 base year variable label
59CA	-	-	Knows 10th grader's 1st friend
59DA	62B	54B	Knows mother of 10th grader's 1st friend
59EA	62B	54B	Knows father of 10th grader's 1st friend
59BB	62A	54A	2nd friend attends same school
59CB	-	-	Knows 10th grader's 2nd friend
59DB	62B	54B	Knows mother of 10th grader's 2nd friend
59EB	62B	54B	Knows father of 10th grader's 2nd friend
59BC	62A	54A	3rd friend attends same school
59CC	-	-	Knows 10th grader's 3rd friend
59DC	62B	54B	Knows mother of 10th grader's 3rd friend
59EC	62B	54B	Knows father of 10th grader's 3rd friend
60A	-	-	Friend's parent gave advice about teachers/courses
60B	-	-	Friend's parent did favor
60C	-	-	Friend's parent received favor
60D	-	-	Friend's parent supervised 10th grader on field trip
61	78	4	10th grader has biological/adoptive parent living outside home
62	-	-	10th grader has contact with non-resident parent
63	79	5	Non-resident parent's participation in education decisions
64A	-	-	Non-resident parent attended school open-house
64B	-	-	Non-resident parent attended PTA/PTO meeting
64C	-	-	Non-resident parent attended parent/teacher conference
64D	-	-	Non-resident parent attended school/class event
65	-	58	Number of years parent has lived in current neighborhood
66	-	59	How involved parent feels in neighborhood/community
67	-	-	Level of crime in neighborhood
68	-	60	How safe is neighborhood
69A	65A	51A	Family rules for 10th grader about maintaining grade average
69B	65B	51B	Family rules for 10th grader about doing homework
69C	65C	-	Family rules for 10th grader about doing household chores
69D	-	-	Family rules for 10th grader about watching TV
70	-	-	Days/week eat at least one meal with 10th grader
71	70	-	Computer in home that 10th grader may use
72	-	-	Computer has access to Internet
73	-	-	Uses computer to communicate with 10th grader's school
74A	-	-	How often e-mails teachers/staff about 10th grader
74B	-	-	How often uses computer to learn about school events
74C	-	-	How often uses computer to express concern over policy
74D	-	-	How often uses computer to select classes for 10th grader
74E	-	-	How often uses computer to get information about homework
75	-	-	School has voice-messaging system
76	-	-	How often use voice-messaging system
77A	-	42D	School assigns too little homework
77B	74C	-	10th grader challenged at school

See notes at end of table.

Table H–4. Intercohort parent questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 2nd follow-up	ELS:2002 base year variable label
77C	74D	-	10th grader working hard at school
77D	-	42H	School preparing students well for jobs in workplace
77E	74H	42G	School preparing students well for college
77F	74I	42I	The school is a safe place
77G	74J	42M	Parents have adequate say in setting school policy
77H	74K	42N	Parents work together supporting school policy
77I	-	-	10th grader's teachers are well trained
77J	-	42P	Drinking on school grounds is problem
77K	-	42Q	Drug use on school grounds is problem
77L	-	42R	Sale/use of drugs on way to/from school is problem
77M	-	42S	Theft on school grounds is problem
77N	-	42T	Violence on school grounds is problem
77O	-	42U	Lack of discipline in class is problem
78	75	41	Satisfaction with 10th grader's education up to now
79	-	61	How far in school wants 10th grader to go
80A	-	66A	Post-sec school's low expenses important to parent
80B	-	66B	Availability of post-sec financial aid important to parent
80C	-	66C	Post-sec school's courses/curriculum important to parent
80D	-	66D	Post-sec school's athletic program important to parent
80E	-	66E	Post-sec school's active social life important to parent
80F	-	66F	Living at home while attending post-sec important to parent
80G	-	66G	Away from home while attending post-sec important to parent
80H	-	66H	Post-sec school's religious environment important to parent
80I	-	66I	Post-sec school's low crime important to parent
80J	-	66J	Post-sec school's job placement record important to parent
80K	-	66K	Post-sec school's grad school placement important to parent
80L	-	66L	Post-sec school's academic reputation important to parent
80M	-	66M	Post-sec school's easy admission important to parent
80N	-	66O	Post-sec school's racial/ethnic makeup important to parent
80O	-	66P	Post-sec school's size important to parent
81	76	-	How far in school parent expects 10th grader will go
82	84	-	Savings efforts for 10th grader's education after high school
83A	84AA	79A	Started a savings account
83B	84AB	79B	Bought an insurance policy
83C	84AC	79C	Bought U.S. savings bonds
83D	84AD	79D	Made investments in stocks/real estate
83E	84AE	79E	Set up a college investment fund
83F	84AF	79F	Started working another job/more hours
83G	84AG	79G	Established another form of savings
83H	-	-	Reduced other expenses in some way
83I	-	79H	Planned to reduce other expenses in some way
83J	-	-	Remortgaged property/took out home-equity loan

See notes at end of table.

Table H–4. Intercohort parent questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 2nd follow-up	ELS:2002 base year variable label
83K	-	79I	Planned to remortgage property/take out home-equity loan
83L	-	79J	Had 10th grader put aside earnings
83M	-	-	Participated in state-sponsored college savings program
84	84B	81	Amount of money set aside for 10th grader's future education
85	80	74	Total family income from all sources 2001
86	81	75	Number of earners contributed to family income
97	-	93	Received help in completing questionnaire
98A	-	94A	10th grader helped with questionnaire
98B	-	94B	Spouse/partner helped with questionnaire
98C	-	94C	Other family member helped with questionnaire
98D	-	94D	A friend helped with questionnaire
98E	-	94E	Other person in community helped with questionnaire
99	-	94	Year/month parent questionnaire completed

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table H–5. Intercohort school administrator questionnaire crosswalk

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
1	2	2	1	Total student enrollment as of October 2001
2A	1A	1A	3A	School has pre-kindergarten
2B	1B	1B	3B	School has kindergarten
2C	1C	1C	3C	School has 1st grade
2D	1D	1D	3D	School has 2nd grade
2E	1E	1E	3E	School has 3rd grade
2F	1F	1F	3F	School has 4th grade
2G	1G	1G	3G	School has 5th grade
2H	1H	1H	3H	School has 6th grade
2I	1I	1I	3I	School has 7th grade
2J	1J	1J	3J	School has 8th grade
2K	1K	1K	3K	School has 9th grade
2L	1L	1L	3L	School has 10th grade
2M	1M	1M	3M	School has 11th grade
2N	1N	1N	3N	School has 12th grade
2O	1O	1O	3O	School has 13th grade or higher
3A	-	4AA	4A	Comprehensive public school
3B	-	4AB	4B	Public magnet school
3C	-	-	-	Public magnet school with theme
3D	-	4AC	4C	Public school of choice
3E	-	4AD	4D	Year round school
3F	-	-	-	Area vocational school/center
3G	-	4AE	-	Full-time technical/vocational school
3H	-	4AE	4F	Other technical or vocational school
3I	4	4AF	4G	Catholic diocesan school
3J	4	4AG	4H	Catholic parish
3K	4	4AH	4I	Catholic religious order
3L	-	-	-	Catholic independent school
3M	4	4AI	4J	Other private school with religious affiliation
3N	4	4AJ	4K	Private school without religious affiliation
3O	-	4AK	4L	Boarding school
3P	-	4AL	4M	Indian reservation school
3Q	-	4AM	4N	Military academy
3R	-	-	4O	Alternative/dropout prevention/continuation school
3S	-	-	-	Charter school
4	-	-	-	Way of teaching students with different abilities
5	-	-	-	Crime in students^ neighborhood
6	-	6	-	Type of academic calendar
7	6	7	5	Number of days in school year for 10th graders
8	7	8	-	Number of class periods in day for 10th graders
9	8	9	-	Number of minutes of average 10th grade class period
10	-	-	6	Typical semester class load for 10th graders

See notes at end of table.

Table H–5. Intercohort school administrator questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
11	-	10	-	School is coeducational
12A	-	13A	-	Percent in school-sponsored community service
12B	-	13B	-	Percent in work study program
12C	-	13C	-	Percent in academic counseling program
12D	-	13D	-	Percent in vocational counseling program
12E	-	13F	-	Percent in dropout prevention program
12F	-	13G	-	Percent in gang prevention program
12G	-	13H	-	Percent in alcohol/drug prevention program
12H	-	13I	-	Percent in AIDS education program
12I	-	13K	-	Percent in crisis prevention program
13	-	21	-	When parents notified of absences
14A	-	11A	7A	Percent 10th graders in general high school program
14B	-	11B	7B	Percent 10th graders in college prep program
14C	-	11C 10	7C	Percent 10th graders in other specialized programs
14D	-	11C 1-9	7D 1-9	Percent 10th graders in voc/tech/business program
14E	-	-	7E	Percent 10th graders in special ed program
14F	-	-	7F	Percent 10th graders in alternative program
14G	-	-	-	Percent 10th graders receive bilingual education
14H	-	-	-	Percent 10th graders receive ESL
14I	-	-	-	Percent 10th graders receive remedial reading
14J	-	-	-	Percent 10th graders receive remedial math
14K	-	-	-	Percent 10th graders in after school/summer outreach
15A	-	-	-	Students develop career plan
15B	-	-	-	Students select career major/pathway
15C	-	-	-	Students in program to prepare for college
16	-	-	16	Vocational-technical programs offered
17A	-	-	-	Agriculture/renewable resource courses offered
17B	-	-	-	Business courses offered
17C	-	-	-	Marketing/distribution courses offered
17D	-	-	-	Health care courses offered
17E	-	-	-	Public/protective service courses offered
17F	-	-	-	Construction courses offered
17G	-	-	-	Mechanics and repair courses offered
17H	-	-	-	Precisions production courses offered
17I	-	-	-	Trade/industry/transportation courses offered
17J	-	-	-	Computer technology courses offered
17K	-	-	-	Communication technology courses offered
17L	-	-	-	Other technology courses offered
17M	-	-	-	Food service and hospitality courses offered
17N	-	-	-	Child care/education courses offered
17O	-	-	-	Personal and other services courses offered
17P	-	-	-	Other occupational courses offered

See notes at end of table.

Table H–5. Intercohort school administrator questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
17Q	-	-	-	Family/consumer sciences courses offered
17R	-	-	-	Industrial arts/technology courses offered
18A	-	-	-	Cooperative education offered to 10th graders
18B	-	-	-	Internships offered to 10th graders
18C	-	-	-	Job shadowing offered to 10th graders
18D	-	-	-	Mentoring offered to 10th graders
18E	-	-	-	Community service offered to 10th graders
18F	-	-	-	School-based enterprise offered to 10th graders
19AA	-	-	-	Baseball offered to males
19AB	-	-	-	Baseball offered to females
19BA	-	-	-	Softball offered to males
19BB	-	-	-	Softball offered to females
19CA	-	-	-	Basketball offered to males
19CB	-	-	-	Basketball offered to females
19DA	-	-	-	Football offered to males
19DB	-	-	-	Football offered to females
19EA	-	-	-	Soccer offered to males
19EB	-	-	-	Soccer offered to females
19FA	-	-	-	Swim team offered to males
19FB	-	-	-	Swim team offered to females
19GA	-	-	-	Ice hockey offered to males
19GB	-	-	-	Ice hockey offered to females
19HA	-	-	-	Field hockey offered to males
19HB	-	-	-	Field hockey offered to females
19IA	-	-	-	Volleyball offered to males
19IB	-	-	-	Volleyball offered to females
19JA	-	-	-	Lacrosse offered to males
19JB	-	-	-	Lacrosse offered to females
19KA	-	-	-	Tennis offered to males
19KB	-	-	-	Tennis offered to females
19LA	-	-	-	Cross-country offered to males
19LB	-	-	-	Cross-country offered to females
19MA	-	-	-	Track offered to males
19MB	-	-	-	Track offered to females
19NA	-	-	-	Golf offered to males
19NB	-	-	-	Golf offered to females
19OA	-	-	-	Gymnastics offered to males
19OB	-	-	-	Gymnastics offered to females
19PA	-	-	-	Wrestling offered to males
19PB	-	-	-	Wrestling offered to females
19QA	-	-	-	Cheerleading offered to males
19QB	-	-	-	Cheerleading offered to females

See notes at end of table.

Table H–5. Intercohort school administrator questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
19RA	-	-	-	Drill team offered to males
19RB	-	-	-	Drill team offered to females
19SA	-	-	-	Other sport offered to males
19SB	-	-	-	Other sport offered to females
19TA	-	-	-	No sports offered to males
19TB	-	-	-	No sports offered to females
20	15	29	24	Percent 10th graders are LEP or non-English proficient
21	16A	30A	25A	Percent 10th graders receive free/reduced-price lunch
22A	17	35	29A	Number of full-time teachers
22B	-	-	29B	Number of part-time teachers
23A	-	41A	36A	Number of full-time math teachers
23B	-	41B	36B	Number of full-time science teachers
23C	-	-	36C	Number of full-time art teachers
23D	-	-	36D	Number of full-time music teachers
23E	-	41D	36E	Number of full-time English teachers
23F	-	41E	36F	Number of full-time foreign language teachers
23G	-	41F	36G	Number of full-time social sciences teachers
23H	-	41G	36H	Number of full-time history teachers
23I	-	41H	36I	Number of full-time vocational education teachers
23J	-	41I	36J	Number of full-time physical education teachers
23K	-	41J	36K	Number full-time guidance counselors
23L	-	41K	36L	Number full-time special education teachers
24A	-	-	-	Percent full-time teachers are certified
24B	-	-	-	Percent part-time teachers are certified
25A	-	-	-	Percent full-time teachers teach out of field
25B	-	-	-	Percent part-time teachers teach out of field
26A	19	42A	37A	Lowest salary paid to full-time teachers
26B	-	42B	37B	Highest salary paid to full-time teachers
27A	-	-	40A	Principal/administrator evaluates teachers
27B	-	-	40B	Teachers evaluate teachers
27C	-	-	40C	Students evaluate teachers
28A	-	53B	41A	Good teachers given special awards
28B	-	53C	41B	Good teachers assigned to better students
28C	-	53E	41C	Good teachers given a lighter teaching load
28D	-	53F	41D	Good teachers relieved of administrative/disciplinary duties
28E	-	53G	41E	Good teachers given priority on requests for materials
28F	-	-	41F	Good teachers receive higher pay
28G	-	53A	-	Good teachers are not recognized in these ways
29	-	-	-	Content standards for academic subjects
30	-	-	-	Main source of content standards
31	-	-	-	Content standards linked with performance standards
32	-	65	42	Students must pass a test for high school diploma

See notes at end of table.

Table H–5. Intercohort school administrator questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
33AA	-	-	43A	Minimum competency test given in grade 7
33AB	-	-	45A	Math is on grade 7 competency test
33AC	-	-	45B	Science is on grade 7 competency test
33AD	-	-	45C	English is on grade 7 competency test
33AE	-	-	45D	History/social studies is on grade 7 competency test
33BA	-	-	43B	Minimum competency test given in grade 8
33BB	-	-	45A	Math is on grade 8 competency test
33BC	-	-	45B	Science is on grade 8 competency test
33BD	-	-	45C	English is on grade 8 competency test
33BE	-	-	45D	History/social studies is on grade 8 competency test
33CA	-	66B	43C	Minimum competency test given in grade 9
33CB	-	-	45A	Math is on grade 9 competency test
33CC	-	-	45B	Science is on grade 9 competency test
33CD	-	-	45C	English is on grade 9 competency test
33CE	-	-	45D	History/social studies is on grade 9 competency test
33DA	-	66C	43D	Minimum competency test given in grade 10
33DB	-	-	45A	Math is on grade 10 competency test
33DC	-	-	45B	Science is on grade 10 competency test
33DD	-	-	45C	English is on grade 10 competency test
33DE	-	-	45D	History/social studies is on grade 10 competency test
33EA	-	66D	43E	Minimum competency test given in grade 11
33EB	-	-	45A	Math is on grade 11 competency test
33EC	-	-	45B	Science is on grade 11 competency test
33ED	-	-	45C	English is on grade 11 competency test
33EE	-	-	45D	History/social studies is on grade 11 competency test
33FA	-	66E	43F	Minimum competency test given in grade 12
33FB	-	-	45A	Math is on grade 12 competency test
33FC	-	-	45B	Science is on grade 12 competency test
33FD	-	-	45C	English is on grade 12 competency test
33FE	-	-	45D	History/social studies is on grade 12 competency test
34A	-	-	44A	Competency test is state requirement
34B	-	-	44B	Competency test is district requirement
34C	-	-	44C	Competency test is school requirement
35	-	-	-	Competency test tied to content standards
36	-	-	46	Percent fail competency test on first attempt
37A	-	-	47A	Retake competency test if failed
37B	-	-	47B	Take remedial class if fail competency test
37C	-	-	47C	Complete competency test preparation class if fail
37D	-	-	47D	Tutoring/individualized academic program if fail competency test
37E	-	-	47E	Summer school if fail competency test
37F	-	-	-	Referred to alternative/continuing ed school if fail competency test
38A	-	-	-	Control access to buildings during school hours

See notes at end of table.

Table H–5. Intercohort school administrator questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
38B	-	-	-	Control access to grounds during school hours
38C	-	-	-	Require students pass through metal detector
38D	-	-	-	Random metal detector checks on students
38E	-	-	-	Close campus for students during lunch
38F	-	-	-	Random dog sniffs to check for drugs
38G	-	-	-	Random sweeps for contraband
38H	-	-	-	Require drug testing for any students
38I	-	-	-	Require students to wear uniforms
38J	-	-	-	Enforce strict dress code
38K	-	-	-	Require clear book bags/ban book bags
38L	-	-	-	Require students to wear badges/picture ID
38M	-	-	-	Require faculty/staff to wear badges/picture ID
38N	-	-	-	Use security cameras to monitor school
38O	-	-	-	Telephones in most classrooms
38P	-	-	-	Emergency call button in classrooms
39A	-	-	-	Process to get parent input on discipline policies
39B	-	-	-	Training parents to deal with problem behavior
39C	-	-	-	Program involves parents in school discipline
40A	-	-	-	Use paid security at any time during school hours
40B	-	-	-	Use paid security as students arrive or leave
40C	-	-	-	Use paid security at school activities
40D	-	-	-	Use paid security outside of school hours/activities
40E	-	-	-	Use paid security at other time
41A	-	-	-	Teachers have access to cable TV
41B	-	-	-	Teachers have access to closed-circuit TV
41C	-	-	-	Teachers have access to videodisc player/VCR/DVD
41D	-	-	-	Teachers have access to video camera
41E	-	-	-	Teachers have access to video production studio
41F	-	-	-	Teachers have access to satellite TV hook-up
41G	-	-	-	Teachers have access to videoconferencing equipment
41H	-	-	-	Teachers have access to digital camera
41I	-	-	-	Teachers have access to scanner
41J	-	-	-	Teachers have access to LCD panel
41K	-	-	-	Teachers have access to laptop computer
41L	-	-	-	Teachers have access to Internet
41M	-	-	-	Teachers have access to computer printer
42A	-	-	-	Teachers use computers as instructional tools
42B	-	-	-	Teachers use computers to plan lessons
42C	-	-	-	Teachers use computers for professional development courses
42D	-	-	-	Teachers use computers to communicate with colleagues
42E	-	-	-	Teachers use computers to access best practices
42F	-	-	-	Teachers use computers to communicate with parents

See notes at end of table.

Table H–5. Intercohort school administrator questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
42G	-	-	-	Teachers use computers to post homework
42H	-	-	-	Teachers/staff use computers to communicate with each other
42I	-	-	-	Teachers use computers to teach job skills
42J	-	-	-	Administrative staff use computers for administrative purposes
42K	-	-	-	Administrative staff use computers to communicate with colleagues
42L	-	-	-	Administrative staff use computers to communicate with parents
42M	-	-	-	School offers students distance learning courses
42N	-	-	-	Teachers have access to Internet professional development programs
43A	-	-	-	Teacher training on use of new software
43B	-	-	-	Teacher training on use of Internet
43C	-	-	-	Teacher training on using computers to teach skills
43D	-	-	-	Teacher training on integrating computer into class
43E	-	-	-	Teacher training on basic computer literacy
44A	-	-	-	Computers in administrative offices
44B	-	-	-	Computers in teacher work rooms
44C	-	-	-	Computers in classrooms
44D	-	-	-	Computers in the library media center
44E	-	-	-	Computers in separate computer lab
46A	-	98A	59A	Principal^s influence on hiring/firing teachers
46B	-	-	52B	Principal^s influence on grouping students
46C	-	-	52C	Principal^s influence on course offerings
46D	-	-	52D	Principal^s influence on instructional materials
46E	-	98E	52E	Principal^s influence on curricular guidelines
46F	-	-	52F	Principal^s influence on grading and evaluation
46G	-	-	52G	Principal^s influence on discipline policies
46H	-	-	59H	Principal^s influence on school funds
47A	-	99C	60C	School^s relationship with school board
47B	-	99D	60D	School^s relationship with central office
47C	-	99E	60E	School^s relationship with teachers^ association
48A	-	103A	62A	Principal evaluated on standardized test scores
48B	-	103B	62B	Principal evaluated on school environment
48C	-	103C	62C	Principal evaluated on efficient administration
48D	-	103D	62D	Principal evaluated on parent involvement
48E	-	103D	62E	Principal evaluated on relationship with community
48F	-	-	62F	Principal evaluated on new programs/reform
49A	49A	95A	57A	How often tardiness a problem at school
49B	49B	95B	57B	How often absenteeism a problem at school
49C	49C	95C	57C	How often class cutting a problem at school
49D	49D	95D	57D	How often physical conflicts a problem at school
49E	49E	95F	57F	How often robbery/theft a problem at school
49F	49F	95G	57G	How often vandalism a problem at school
49G	49G	95H	57H	How often use of alcohol a problem at school

See notes at end of table.

Table H–5. Intercohort school administrator questionnaire crosswalk–Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
49H	49H	95I	57I	How often use of illegal drugs a problem at school
49I	-	-	57J	How often students on drugs/alcohol at school a problem
49J	-	-	57K	How often sale of drugs near school a problem
49K	49I	95J	57L	How often possession of weapons a problem at school
49L	49J	95K	57M	How often physical abuse of teachers a problem at school
49M	-	95M	57O	How often racial tension among students a problem at school
49N	-	-	-	How often student bullying a problem at school
49O	49K	95I	57N	How often verbal abuse of teachers a problem at school
49P	-	-	-	How often disorder in classrooms a problem at school
49Q	-	-	-	How often student disrespect for teachers a problem at school
49R	-	95E	57E	How often gang activity a problem at school
49S	-	-	-	How often cult/extremist group activities a problem at school
50A	-	-	-	Learning hindered by poor condition of buildings
50B	-	-	-	Learning hindered by poor heating/air/light
50C	-	-	-	Learning hindered by poor science labs
50D	-	-	-	Learning hindered by poor fine arts facilities
50E	-	-	-	Learning hindered by lack of space
50F	-	-	-	Learning hindered by poor library
50G	-	-	-	Learning hindered by lack of texts/supplies
50H	-	-	-	Learning hindered by too few computers
50I	-	-	-	Learning hindered by lack of multi-media
50J	-	-	-	Learning hindered by lack of discipline/safety
50K	-	-	-	Learning hindered by poor voc/tech equipment/facilities
51A	-	-	-	Student morale is high
51B	-	-	-	Teachers press students to achieve
51C	-	-	-	Teacher morale is high
51D	-	-	-	Learning is high priority for students
51E	-	-	-	Students expected to do homework
53	-	104	63	Date completed interview

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

Table H–6. Intercohort teacher questionnaire crosswalk

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
1	I-1	1	I-1A	Taught student in fall 2001
2	.	.	.	How well remembers student from fall semester
3	I-1	.	I-1B	Taught student in spring 2002
4	.	I-2	I-2	Student usually works hard for good grades
5	.	I-3	I-3	Student relates well to others
6	I-7	I-19	.	Student is exceptionally passive
7	.	I-5	I-5	Student talks with teacher outside of class
8A	.	I-6A	I-6A	Spoke to parents about poor performance
8B	.	I-6B	I-6B	Spoke to parents about disruptive behavior
8C	.	6C	6C	Spoke to parents about not doing homework
8D	.	6 D	6D	Spoke to parents about absenteeism
8E	.	.	.	Spoke to parents about accomplishments
9	.	I-7	I-7	Parents' level of involvement in academic performance
10	.	I-13	I 8	Difficulty of class for student
11	I-10	I-9,10	.	Student has disability that affects school work
12	.	.	.	Student has fallen behind in school work
12A	I-9	I-8	.	Student behind due to health problem
12B	.	.	.	Student behind due to LEP
12C	.	.	.	Student behind due to disciplinary action
12D	.	.	.	Student behind due to lack of effort
12E	.	.	.	Student behind due to other reason
13	I-3	I-15	I-12	How often student completes homework
14	I-4	I-16	I-13	How often student is absent
15	I-5	I-17	I-14	How often student is tardy
16	I-6	I-18	I-15	How often student is attentive in class
17	I-8	I-20	I-16	How often student is disruptive in class
18A	.	I-21A	I-17A	Spoke to counselor about poor performance
18B	.	I-21B	I-17B	Spoke to counselor about disruptive behavior
19	.	I-14	.	Recommended student for AP/honors classes/academic honors
20	.	.	.	How far teacher expects student to get in school
21A	.	.	.	How well student organizes ideas
21B	.	.	.	How well student uses grammar
21C	.	.	.	How well student uses appropriate detail
21D	.	.	.	How well student expresses critical/creative thought
22	III-1	III-1	IV-1	Teacher's sex
23	.	.	.	Teacher is Hispanic
24A	III-2	III-2	IV-2	Teacher is White
24B	III-2	III-2	IV-2	Teacher is Black/African American
24C	III-2	III-2	IV-2	Teacher is Asian
24D	III-2	III-2	IV-2	Teacher is Native Hawaiian/Pacific Islander
24E	III-2	III-2	IV-2	Teacher is American Indian/Alaska Native
25	III-3C	I-3C	IV-3	Teacher's year of birth
26A	.	III-4A	IV-4A	Years teaching at elementary level/K-6
26B	.	III-4B	IV-4B	Years teaching at secondary level/7-12
26C	III-4	.	.	Total years teaching/K-12
27	III-5	III-5	IV-5	Total years teaching in this school
28	III-18	III-6	IV-6	Employment status in this school/system
29	III-6	III-7	IV-7A	Type of certification held

See notes at end of table.

Table H–6. Intercohort teacher questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
30A	III-8	III-9A	III-8A	No academic degree held
30B	III-8	III-9B	III-8B	Associate degree held
30C	III-8	III-9C	III-8C	Bachelor's degree held
30D	III-8	III-9D	III-8E	Education specialist degree held
30E	III-8	III-9E	III-8D	Master's degree held
30F	III-8	III-9F	III-8F	Doctorate degree held
30G	III-8	III-9G	III-8G	First professional degree held
31A	III-9A	III-10A	IV-9A	Bachelor's degree major
31B	III-9B	III-10B	IV-9B	Bachelor's degree minor/2nd major
32A	10A	III-11A	IV-10A	Highest graduate degree major
32B	10B	III-11B	IV-10B	Highest graduate degree minor/2nd major
33A	.	III-14A	IV-13A	Number undergraduate English courses taken
33B	.	III-14B	IV-13B	Number graduate English courses taken
33C	.	III-14A	IV-13A	Number undergraduate math courses taken
33D	.	III-14B	IV-13B	Number graduate math courses taken
34	.	III-16	.	If starting over whether would be a teacher again
35A	.	.	.	How often use computer to create materials
35B	.	.	.	How often use WWW sites to plan lessons
35C	.	.	.	How often access model lesson plans from Internet
35D	.	.	.	How often research teaching on Internet
35E	.	.	.	How often take professional development courses on Internet
35F	.	.	.	How often use Internet for colleague discussions
35G	.	.	.	How often download instructional software from Internet
35H	.	.	.	How often use computer to give class presentations
35I	.	.	.	How often use computer for administrative records
35J	.	.	.	How often use computer to prepare multimedia presentations
35K	.	.	.	How often use computer to communicate w/colleagues
35L	.	.	.	How often use computer to communicate w/parents
35M	.	.	.	How often use computer to communicate w/students
35N	.	.	.	How often use computer to post homework/information
36	.	.	.	Hours of training on teaching special education students
37	.	.	.	Has had eight hours training on teaching LEP students
38A	.	.	.	Received training in basic computer skills
38B	.	.	.	Received training in software applications
38C	.	.	.	Received training in use of Internet
38D	.	.	.	Received training in use of other technology
38E	.	.	.	Received training in integrating technology in curriculum
38F	.	.	.	Received follow-up or advanced training
39	III-28	III-21	IV-22 A-B	Days missed teaching during 1st semester
40	III-27	III-17A	.	Holds additional full-time job
41	.	III-18	.	Additional full-time job related to education
42	III-27	III-17B	.	Holds additional part-time job
43	.	III-18	.	Additional part-time job related to education
44A	.	IV-4	.	Importance of home background to student success
44B	.	IV-4	.	Importance of intellectual ability to student success
44C	.	IV-4	.	Importance of student's enthusiasm to student success
44D	.	IV-4	.	Importance of teacher's attention to student success
44E	.	IV-4	.	Importance of teaching methods to student success

See notes at end of table.

Table H–6. Intercohort teacher questionnaire crosswalk—Continued

ELS:2002 base year	NELS:88 base year	NELS:88 1st follow-up	NELS:88 2nd follow-up	ELS:2002 base year variable label
44F	.	IV-4	.	Importance of teacher's enthusiasm to student success
45A	.	.	.	People can learn to be good at math
45B	.	.	.	People must be born with math ability
47	.	IV-12	IV-23	Date teacher questionnaire completed

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

H.2 Reference

Nakao, K., and Treas, J. (1992). *The 1989 Socioeconomic Index of Occupations: Construction from the 1989 Occupational Prestige Scores*. General Social Survey Methodological Report No. 74. Chicago: National Opinion Research Center.

Appendix I
Item Nonresponse Bias Analysis Tables

This page is intentionally blank.

Appendix I
Item Nonresponse Bias Analysis Tables

Web-published PDF files of item nonresponse bias analysis tables are available at: http://nces.ed.gov/surveys/els2002/

This page is intentionally blank.

Appendix J
Details of School and Student Sampling

This page is intentionally blank.

Appendix J
Details of School and Student Sampling

J.1 School Sampling

This section gives the mathematical details of the school sampling design for the full-scale survey. A composite measure of size sampling approach was used to select the school sample because, as demonstrated by Folsom et al. (1987), composite measure of size sampling designs are useful for achieving self-weighting samples for multiple analysis domains (e.g., student by school strata) in multistage sampling designs with equal workloads for all primary sampling units (schools).

Defining notation for the strata, the student sampling rates, and the composite measure of size for schools begins as follows:

(1) r = 1, 2, ..., 66 indexes the school strata (region by metro status by public/Catholic/other private).

(2) s = 1, 2, 3 indexes the student strata.

(3) i = 1, 2, ..., I(r) indexes the schools in stratum "r."

(4) $M_{rs}(i)$ = number of students enrolled in the 10th grade in 2002 who belonged to student stratum "s" at the i-th school in stratum "r" based on the latest Common Core of Data (CCD) and Private School Survey (PSS) data.

(5) m_{rs} = number of students, adjusted for nonresponse, selected from student stratum "s" within the r-th school stratum, referred to henceforth as student stratum "rs."

The overall population sampling rate for student stratum "rs" then is given by

$$f_{rs} = m_{rs} / M_{rs}(+) \ ,$$

where

$$M_{rs}(+) = \sum_{i=1}^{I(r)} M_{rs}(i) \quad .$$

The student sampling rates, f_{rs}, were computed based on the final sample allocation and frame data regarding the population sizes.

The composite measure of size for the i-th school in stratum "r" was then defined as was then defined as

$$S_r(i) = \sum_{s=1}^{3} f_{rs} M_{rs}(i) \quad ,$$

which is the number of students who would be selected from the i-th school if all schools on the frame were to be sampled.

An independent sample of schools was selected for each school stratum using Chromy's sequential probability with minimum replacement (pmr) sampling algorithm to select schools with probabilities proportional to their measures of size (Chromy 1979). There were no schools with an expected frequency of selection greater than unity (1.00). Therefore, the expected frequency of selection for the i-th school in school stratum "r" was given by

$$\pi_r(i) = \frac{n_r^* S_r(i)}{S_r(+)}, \text{for non-certainty selections;}$$

where

$$S_r(+) = \sum_{i=1}^{I(r)} S_r(i),$$

and n_r is the number of non-certainty selections from stratum "r."
Within each of the "r" school strata, implicit stratification was achieved by sorting the stratum "r" sampling frame in a serpentine manner (see Williams and Chromy 1980) by state. The objectives of this additional, implicit stratification were to ensure proportionate representation of all states.

J.2 Student Sampling

Recall that the overall population sampling rate for student stratum "rs" was given by

$$f_{rs} = m_{rs} / M_{rs}(+) \quad ,$$

where

$$M_{rs}(+) = \sum_{i=1}^{I(r)} M_{rs}(i) \quad .$$

For the unconditional probability of selection to be a constant for all eligible students in stratum "rs," the overall probability of selection should be the overall student sampling fraction, f_{rs}; i.e., it was required that

$$\frac{m_{rs}(i)}{M_{rs}(i)} \pi_r(i) = f_{rs} \quad ,$$

or equivalently,

$$m_{rs}(i) = f_{rs} \frac{M_{rs}(i)}{\pi_r(i)} \quad .$$

Thus, the conditional sampling rate for stratum "rs," given selection of the i-th school, became

$$f_{rs\backslash i} = f_{rs} / \pi_r(i)$$

However, in this case, the desired overall student sample size, ms, was achieved only in expectation over all possible samples.

Achieving the desired sample sizes with equal probabilities within strata in the particular sample that has been selected, and simultaneously adjusting for school nonresponse and ineligibility, required that

$$\sum_{i \varepsilon R} m_{rs}(i) = m_{rs} \quad ,$$

where "R" denoted the set of eligible, responding schools. If the conditional student sampling rate for stratum "rs" in the i-th school was

$$\hat{f}_{rs \backslash i} = \hat{f}_{rs} / \pi_r(i) \quad ,$$

it then required

$$\sum_{i \varepsilon R} \hat{f}_{rs} \frac{M_{rs}(i)}{\pi_r(i)} = m_{rs} \quad ,$$

or equivalently,

$$\hat{f}_{rs} = m_{rs} / \hat{M}_{rs} \quad ,$$

where

$$\hat{M}_{rs} = \frac{\sum_{i \varepsilon R} M_{rs}(i)}{\pi_r(i)} \quad .$$

Since it was necessary to set the student sampling rates before having complete information on eligibility and response status, M_{rs} was calculated as follows:

$$\hat{M}_{rs} = \sum_{i \varepsilon S} \frac{M_{rs}(i)}{\pi_r(i)} * [E_r R_r E_{rs}] \quad ,$$

where "S" denotes the set of all sample schools,

E_r = the school eligibility factor for school stratum "r,"
R_r = the school response factor for school stratum "r,"
E_{rs} = the student eligibility factor for student stratum "rs."

J.3 References

Folsom, R.E., Potter, F.J., and Williams, S.R. (1987). Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains. *Proceedings of the Survey Research Methods Section* (pp. 792–796). The American Statistical Association.

Chromy, J.R. (1979). Sequential Sample Selection Methods. *Proceedings of the Survey Research Methods Section* (pp. 401–406). The American Statistical Association.

Williams, R.L., and Chromy, J.R. (1980). SAS Sample Selection MACROS. *Proceedings of the Fifth Annual SAS Users Group International Conference* (pp. 392–396).

This page is intentionally blank.

Appendix K
Standard Errors and Design Effects

This page is intentionally blank.

Table K–1. Parent design effects – all

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	92.1	0.28	0.23	13487	1.48	1.22
Student lives with you all of the time	BYP05	94.6	0.24	0.20	13440	1.52	1.23
Average number of people dependent on parent or spouse/partner	BYP06	2.7	0.02	0.01	12505	2.24	1.50
Married	BYP10	72.7	0.51	0.38	13422	1.76	1.33
Biological mother born in the U.S.	BYP17	81.9	0.68	0.33	13391	4.18	2.04
Parent's native language is English	BYP28	85.0	0.68	0.31	13361	4.80	2.19
Parent attended college, no 4-year degree	PARED	11.9	0.32	0.26	15362	1.46	1.21
Parent works full-time	BYP36	63.1	0.53	0.42	13378	1.61	1.27
Mother works in a professional occupation (groups a or b)	OCCUMOTH	17.3	0.46	0.30	15362	2.29	1.51
Spouse/partner works full-time	BYP40	82.0	0.53	0.37	10512	1.97	1.40
Father works in a professional occupation (groups a or b)	OCCUFATH	14.8	0.44	0.29	15362	2.37	1.54
Child changed school two times since first grade	BYP45	12.1	0.36	0.29	12394	1.48	1.22
Child was held back a grade	BYP46	12.6	0.40	0.30	12430	1.79	1.34
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	35.2	0.65	0.43	12308	2.28	1.51
Belong to school's parent-teacher organization	BYP54A	24.2	0.74	0.39	12271	3.67	1.92
Always discuss child's report card with child	BYP55B	86.7	0.37	0.30	12404	1.49	1.22
Provided advice or information about applying to college or other schools after high school	BYP56C	73.8	0.50	0.40	12310	1.59	1.26
Parent sometimes or frequently attended school activities	BYP57A	69.4	0.56	0.41	12428	1.87	1.37
Parent of child's friend did me a favor	BYP60B	65.2	0.60	0.43	12176	1.96	1.40
Child has another parent living outside of home	BYP61	32.1	0.54	0.42	12316	1.65	1.28
Average number of years living in current neighborhood	BYP65	10.7	0.12	0.08	12418	2.23	1.49
Low level of crime in neighborhood	BYP67	88.6	0.47	0.29	12323	2.72	1.65
Family rules enforced for doing homework	BYP69B	92.9	0.28	0.23	12320	1.48	1.22
Parent and child eat at least one meal together three times a week	BYP70	8.8	0.32	0.24	13316	1.70	1.31
Computer in home for 10th grader to use	BYP71	86.8	0.48	0.29	13376	2.71	1.65
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	51.0	1.33	0.74	4595	3.24	1.80
Child's school has a voice-messaging system	BYP75	54.1	1.02	0.45	12392	5.17	2.27
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.09	0.08	15362	1.31	1.14
Very important that child's school after high school is in a low crime environment	BYP80I	88.5	0.37	0.29	11786	1.60	1.26
Started a savings account for child's education after high school	BYP83A	76.1	0.70	0.55	6056	1.61	1.27
SUMMARY STATISTICS							
Mean						2.24	1.47
Minimum						1.31	1.14
Median						1.83	1.35
Maximum						5.17	2.27
Standard deviation						1.01	0.30

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–2. Parent design effects – male

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	91.7	0.37	0.34	6692	1.20	1.09
Student lives with you all of the time	BYP05	94.5	0.33	0.28	6669	1.37	1.17
Average number of people dependent on parent or spouse/partner	BYP06	2.6	0.02	0.02	6187	1.57	1.25
Married	BYP10	73.0	0.71	0.54	6664	1.70	1.30
Biological mother born in the U.S.	BYP17	82.2	0.75	0.47	6640	2.56	1.60
Parent's native language is English	BYP28	85.8	0.70	0.43	6627	2.67	1.63
Parent attended college, no 4-year degree	PARED	12.0	0.45	0.37	7646	1.49	1.22
Parent works full-time	BYP36	64.0	0.72	0.59	6638	1.51	1.23
Mother works in a professional occupation (groups a or b)	OCCUMOTH	18.0	0.59	0.44	7646	1.79	1.34
Spouse/partner works full-time	BYP40	81.2	0.76	0.54	5255	1.98	1.41
Father works in a professional occupation (groups a or b)	OCCUFATH	14.5	0.54	0.40	7646	1.79	1.34
Child changed school two times since first grade	BYP45	12.0	0.53	0.42	6126	1.61	1.27
Child was held back a grade	BYP46	15.4	0.60	0.46	6143	1.67	1.29
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	32.6	0.82	0.60	6063	1.86	1.36
Belong to school's parent-teacher organization	BYP54A	23.6	0.84	0.55	6062	2.37	1.54
Always discuss child's report card with child	BYP55B	87.4	0.53	0.42	6130	1.54	1.24
Provided advice or information about applying to college or other schools after high school	BYP56C	72.5	0.76	0.57	6072	1.74	1.32
Parent sometimes or frequently attended school activities	BYP57A	67.2	0.76	0.60	6133	1.62	1.27
Parent of child's friend did me a favor	BYP60B	64.0	0.76	0.62	6010	1.49	1.22
Child has another parent living outside of home	BYP61	31.3	0.71	0.59	6086	1.44	1.20
Average number of years living in current neighborhood	BYP65	10.9	0.16	0.11	6133	1.97	1.40
Low level of crime in neighborhood	BYP67	88.2	0.61	0.41	6085	2.18	1.48
Family rules enforced for doing homework	BYP69B	92.9	0.39	0.33	6090	1.41	1.19
Parent and child eat at least one meal together three times a week	BYP70	8.3	0.42	0.34	6609	1.51	1.23
Computer in home for 10th grader to use	BYP71	86.8	0.59	0.42	6635	2.04	1.43
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	53.7	1.53	1.03	2352	2.20	1.48
Child's school has a voice-messaging system	BYP75	54.2	1.15	0.64	6133	3.27	1.81
Want child to attend college, no 4-year degree	PARASPIR	1.2	0.15	0.13	7646	1.48	1.21
Very important that child's school after high school is in a low crime environment	BYP80I	86.8	0.55	0.44	5793	1.51	1.23
Started a savings account for child's education after high school	BYP83A	76.3	0.87	0.78	2992	1.26	1.12
SUMMARY STATISTICS							
Mean						1.79	1.33
Minimum						1.20	1.09
Median						1.65	1.28
Maximum						3.27	1.81
Standard deviation						0.46	0.16

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–3. Parent design effects – female

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	92.5	0.39	0.32	6795	1.52	1.23
Student lives with you all of the time	BYP05	94.7	0.36	0.27	6771	1.75	1.32
Average number of people dependent on parent or spouse/partner	BYP06	2.7	0.02	0.02	6318	1.75	1.32
Married	BYP10	72.3	0.69	0.54	6758	1.60	1.26
Biological mother born in the U.S.	BYP17	81.7	0.86	0.47	6751	3.30	1.82
Parent's native language is English	BYP28	84.2	0.87	0.44	6734	3.86	1.96
Parent attended college, no 4-year degree	PARED	11.8	0.46	0.37	7716	1.60	1.26
Parent works full-time	BYP36	62.1	0.74	0.59	6740	1.59	1.26
Mother works in a professional occupation (groups a or b)	OCCUMOTH	16.6	0.58	0.42	7716	1.90	1.38
Spouse/partner works full-time	BYP40	82.8	0.65	0.52	5257	1.57	1.25
Father works in a professional occupation (groups a or b)	OCCUFATH	15.0	0.58	0.41	7716	2.07	1.44
Child changed school two times since first grade	BYP45	12.2	0.52	0.41	6268	1.58	1.26
Child was held back a grade	BYP46	9.9	0.45	0.38	6287	1.41	1.19
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	37.8	0.82	0.61	6245	1.78	1.33
Belong to school's parent-teacher organization	BYP54A	24.8	0.90	0.55	6209	2.68	1.64
Always discuss child's report card with child	BYP55B	86.1	0.53	0.44	6274	1.46	1.21
Provided advice or information about applying to college or other schools after high school	BYP56C	75.0	0.65	0.55	6238	1.38	1.18
Parent sometimes or frequently attended school activities	BYP57A	71.6	0.76	0.57	6295	1.80	1.34
Parent of child's friend did me a favor	BYP60B	66.5	0.80	0.60	6166	1.75	1.32
Child has another parent living outside of home	BYP61	32.8	0.78	0.60	6230	1.71	1.31
Average number of years living in current neighborhood	BYP65	10.5	0.14	0.11	6285	1.69	1.30
Low level of crime in neighborhood	BYP67	88.9	0.55	0.40	6238	1.90	1.38
Family rules enforced for doing homework	BYP69B	92.8	0.39	0.33	6230	1.39	1.18
Parent and child eat at least one meal together three times a week	BYP70	9.2	0.47	0.35	6707	1.79	1.34
Computer in home for 10th grader to use	BYP71	86.9	0.61	0.41	6741	2.24	1.50
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	48.2	1.67	1.06	2243	2.52	1.59
Child's school has a voice-messaging system	BYP75	54.0	1.15	0.63	6259	3.32	1.82
Want child to attend college, no 4-year degree	PARASPIR	0.8	0.12	0.10	7716	1.26	1.12
Very important that child's school after high school is in a low crime environment	BYP80I	90.2	0.49	0.38	5993	1.63	1.28
Started a savings account for child's education after high school	BYP83A	76.0	1.04	0.77	3064	1.83	1.35
SUMMARY STATISTICS							
Mean						1.92	1.37
Minimum						1.26	1.12
Median						1.75	1.32
Maximum						3.86	1.96
Standard deviation						0.62	0.20

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–4. Parent design effects – Indian

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	86.5	4.58	3.27	110	1.96	1.40
Student lives with you all of the time	BYP05	90.1	3.04	2.87	109	1.12	1.06
Average number of people dependent on parent or spouse/partner	BYP06	2.9	0.23	0.16	103	2.04	1.43
Married	BYP10	55.2	6.98	4.79	109	2.13	1.46
Biological mother born in the U.S.	BYP17	91.9	3.13	2.62	109	1.43	1.20
Parent's native language is English	BYP28	84.8	5.36	3.44	110	2.43	1.56
Parent attended college, no 4-year degree	PARED	18.7	3.67	3.42	131	1.15	1.07
Parent works full-time	BYP36	65.2	4.46	4.58	109	0.95	0.97
Mother works in a professional occupation (groups a or b)	OCCUMOTH	22.2	3.54	3.64	131	0.95	0.97
Spouse/partner works full-time	BYP40	82.2	4.96	4.51	73	1.21	1.10
Father works in a professional occupation (groups a or b)	OCCUFATH	16.5	5.40	3.25	131	2.75	1.66
Child changed school two times since first grade	BYP45	15.4	4.59	3.61	101	1.62	1.27
Child was held back a grade	BYP46	17.7	3.93	3.80	102	1.07	1.03
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	26.2	4.72	4.40	101	1.15	1.07
Belong to school's parent-teacher organization	BYP54A	10.4	3.88	3.05	101	1.62	1.27
Always discuss child's report card with child	BYP55B	81.2	4.63	3.91	101	1.40	1.18
Provided advice or information about applying to college or other schools after high school	BYP56C	64.1	5.94	4.80	101	1.54	1.24
Parent sometimes or frequently attended school activities	BYP57A	69.0	5.41	4.65	100	1.35	1.16
Parent of child's friend did me a favor	BYP60B	64.4	6.15	4.79	101	1.65	1.28
Child has another parent living outside of home	BYP61	46.5	6.72	5.04	99	1.78	1.33
Average number of years living in current neighborhood	BYP65	11.9	1.19	1.24	98	0.92	0.96
Low level of crime in neighborhood	BYP67	80.1	6.47	4.01	100	2.59	1.61
Family rules enforced for doing homework	BYP69B	93.5	2.79	2.49	99	1.25	1.12
Parent and child eat at least one meal together three times a week	BYP70	7.0	3.22	2.49	106	1.67	1.29
Computer in home for 10th grader to use	BYP71	68.5	5.67	4.49	108	1.59	1.26
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	53.1	13.61	11.16	21	1.49	1.22
Child's school has a voice-messaging system	BYP75	44.4	5.93	4.97	101	1.42	1.19
Want child to attend college, no 4-year degree	PARASPIR	1.6	1.03	1.10	131	0.89	0.94
Very important that child's school after high school is in a low crime environment	BYP80I	88.2	3.61	3.42	90	1.12	1.06
Started a savings account for child's education after high school	BYP83A	84.8	4.96	5.83	39	0.72	0.85
SUMMARY STATISTICS							
Mean							1.21
Minimum							0.85
Median					1.50		1.19
Maximum					0.72		1.66
Standard deviation					1.43		0.20
					2.75		
					0.51		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–5. Parent design effects – Asian

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	83.7	1.32	1.04	1274	1.63	1.28
Student lives with you all of the time	BYP05	95.8	0.56	0.56	1271	1.00	1.00
Average number of people dependent on parent or spouse/partner	BYP06	2.9	0.08	0.05	1022	2.04	1.43
Married	BYP10	82.6	1.16	1.07	1265	1.18	1.08
Biological mother born in the U.S.	BYP17	13.3	1.82	0.96	1251	3.59	1.90
Parent's native language is English	BYP28	26.2	2.19	1.24	1268	3.13	1.77
Parent attended college, no 4-year degree	PARED	7.4	0.87	0.69	1465	1.61	1.27
Parent works full-time	BYP36	63.5	1.93	1.36	1260	2.03	1.42
Mother works in a professional occupation (groups a or b)	OCCUMOTH	16.7	1.50	0.97	1465	2.37	1.54
Spouse/partner works full-time	BYP40	65.3	2.29	1.46	1057	2.44	1.56
Father works in a professional occupation (groups a or b)	OCCUFATH	21.3	1.74	1.07	1465	2.66	1.63
Child changed school two times since first grade	BYP45	13.4	1.28	1.07	1014	1.42	1.19
Child was held back a grade	BYP46	8.5	1.02	0.88	1019	1.36	1.17
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	32.4	2.26	1.48	1003	2.34	1.53
Belong to school's parent-teacher organization	BYP54A	23.7	2.08	1.34	1002	2.40	1.55
Always discuss child's report card with child	BYP55B	68.4	1.81	1.46	1012	1.54	1.24
Provided advice or information about applying to college or other schools after high school	BYP56C	76.0	1.91	1.35	996	1.99	1.41
Parent sometimes or frequently attended school activities	BYP57A	55.0	2.23	1.56	1018	2.04	1.43
Parent of child's friend did me a favor	BYP60B	53.8	2.14	1.61	961	1.77	1.33
Child has another parent living outside of home	BYP61	11.8	1.34	1.02	1011	1.75	1.32
Average number of years living in current neighborhood	BYP65	8.4	0.32	0.22	1005	2.10	1.45
Low level of crime in neighborhood	BYP67	85.4	1.52	1.12	996	1.85	1.36
Family rules enforced for doing homework	BYP69B	90.3	1.18	0.94	992	1.59	1.26
Parent and child eat at least one meal together three times a week	BYP70	6.4	0.91	0.69	1251	1.71	1.31
Computer in home for 10th grader to use	BYP71	92.2	1.22	0.75	1257	2.60	1.61
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	47.5	3.88	2.64	360	2.17	1.47
Child's school has a voice-messaging system	BYP75	55.1	2.24	1.56	1015	2.06	1.44
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.27	0.26	1465	1.10	1.05
Very important that child's school after high school is in a low crime environment	BYP80I	83.0	1.67	1.21	958	1.89	1.37
Started a savings account for child's education after high school	BYP83A	75.1	2.43	1.86	540	1.70	1.30
SUMMARY STATISTICS							
Mean						1.97	1.39
Minimum						1.00	1.00
Median						1.94	1.39
Maximum						3.59	1.90
Standard deviation						0.57	0.20

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–6. Parent design effects – Black

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	87.7	0.88	0.80	1703	1.21	1.10
Student lives with you all of the time	BYP05	95.0	0.63	0.53	1698	1.44	1.20
Average number of people dependent on parent or spouse/partner	BYP06	2.8	0.05	0.04	1472	1.46	1.21
Married	BYP10	48.0	1.25	1.22	1689	1.06	1.03
Biological mother born in the U.S.	BYP17	89.7	1.16	0.74	1689	2.44	1.56
Parent's native language is English	BYP28	96.1	0.59	0.47	1684	1.60	1.26
Parent attended college, no 4-year degree	PARED	14.5	0.92	0.78	2033	1.38	1.17
Parent works full-time	BYP36	68.7	1.49	1.13	1681	1.73	1.31
Mother works in a professional occupation (groups a or b)	OCCUMOTH	13.9	0.86	0.77	2033	1.27	1.13
Spouse/partner works full-time	BYP40	78.9	1.63	1.35	912	1.45	1.20
Father works in a professional occupation (groups a or b)	OCCUFATH	9.3	0.67	0.65	2033	1.06	1.03
Child changed school two times since first grade	BYP45	13.9	1.06	0.91	1458	1.38	1.17
Child was held back a grade	BYP46	21.8	1.28	1.07	1477	1.42	1.19
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	31.3	1.41	1.22	1445	1.34	1.16
Belong to school's parent-teacher organization	BYP54A	21.9	1.31	1.09	1439	1.44	1.20
Always discuss child's report card with child	BYP55B	87.9	1.07	0.86	1458	1.56	1.25
Provided advice or information about applying to college or other schools after high school	BYP56C	79.0	1.28	1.08	1433	1.42	1.19
Parent sometimes or frequently attended school activities	BYP57A	67.6	1.36	1.23	1454	1.23	1.11
Parent of child's friend did me a favor	BYP60B	57.3	1.80	1.31	1419	1.88	1.37
Child has another parent living outside of home	BYP61	45.7	1.52	1.32	1434	1.33	1.15
Average number of years living in current neighborhood	BYP65	10.3	0.35	0.26	1456	1.75	1.32
Low level of crime in neighborhood	BYP67	79.4	1.24	1.06	1446	1.37	1.17
Family rules enforced for doing homework	BYP69B	95.9	0.61	0.52	1452	1.36	1.17
Parent and child eat at least one meal together three times a week	BYP70	13.0	0.97	0.82	1665	1.39	1.18
Computer in home for 10th grader to use	BYP71	71.5	1.47	1.10	1674	1.77	1.33
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	42.5	2.99	2.32	456	1.66	1.29
Child's school has a voice-messaging system	BYP75	56.5	1.88	1.30	1448	2.08	1.44
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.24	0.23	2033	1.12	1.06
Very important that child's school after high school is in a low crime environment	BYP80I	92.2	0.86	0.72	1378	1.44	1.20
Started a savings account for child's education after high school	BYP83A	81.1	1.72	1.54	646	1.24	1.11
SUMMARY STATISTICS							
Mean						1.48	1.21
Minimum						1.06	1.03
Median						1.42	1.19
Maximum						2.44	1.56
Standard deviation						0.30	0.12

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–7. Parent design effects – Hispanic

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	89.9	0.83	0.69	1924	1.44	1.20
Student lives with you all of the time	BYP05	94.1	0.70	0.54	1921	1.69	1.30
Average number of people dependent on parent or spouse/partner	BYP06	3.1	0.05	0.04	1821	2.19	1.48
Married	BYP10	70.2	1.22	1.05	1915	1.37	1.17
Biological mother born in the U.S.	BYP17	40.7	1.98	1.13	1902	3.08	1.75
Parent's native language is English	BYP28	41.1	1.93	1.13	1913	2.95	1.72
Parent attended college, no 4-year degree	PARED	11.5	0.87	0.68	2234	1.68	1.29
Parent works full-time	BYP36	57.0	1.44	1.14	1901	1.61	1.27
Mother works in a professional occupation (groups a or b)	OCCUMOTH	10.1	0.75	0.64	2234	1.37	1.17
Spouse/partner works full-time	BYP40	75.1	1.39	1.12	1483	1.54	1.24
Father works in a professional occupation (groups a or b)	OCCUFATH	7.6	0.67	0.56	2234	1.44	1.20
Child changed school two times since first grade	BYP45	15.6	1.02	0.85	1805	1.43	1.20
Child was held back a grade	BYP46	16.2	1.00	0.87	1809	1.33	1.15
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	27.4	1.25	1.06	1775	1.39	1.18
Belong to school's parent-teacher organization	BYP54A	12.2	0.90	0.77	1789	1.35	1.16
Always discuss child's report card with child	BYP55B	84.3	1.04	0.86	1810	1.49	1.22
Provided advice or information about applying to college or other schools after high school	BYP56C	70.5	1.23	1.08	1782	1.29	1.13
Parent sometimes or frequently attended school activities	BYP57A	55.3	1.41	1.17	1809	1.45	1.20
Parent of child's friend did me a favor	BYP60B	48.9	1.50	1.19	1760	1.59	1.26
Child has another parent living outside of home	BYP61	28.0	1.38	1.06	1790	1.69	1.30
Average number of years living in current neighborhood	BYP65	8.7	0.24	0.19	1815	1.64	1.28
Low level of crime in neighborhood	BYP67	81.7	1.35	0.91	1792	2.18	1.48
Family rules enforced for doing homework	BYP69B	92.9	0.66	0.61	1803	1.19	1.09
Parent and child eat at least one meal together three times a week	BYP70	6.9	0.59	0.58	1903	1.02	1.01
Computer in home for 10th grader to use	BYP71	74.3	1.47	1.00	1910	2.17	1.47
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	44.5	3.26	2.40	429	1.84	1.36
Child's school has a voice-messaging system	BYP75	57.1	1.44	1.16	1815	1.54	1.24
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.25	0.21	2234	1.38	1.17
Very important that child's school after high school is in a low crime environment	BYP80I	90.5	0.88	0.72	1672	1.52	1.23
Started a savings account for child's education after high school	BYP83A	74.1	2.38	1.77	611	1.81	1.35
SUMMARY STATISTICS							
Mean						1.66	1.28
Minimum						1.02	1.01
Median						1.53	1.24
Maximum						3.08	1.75
Standard deviation						0.46	0.16

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–8. Parent design effects – White

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	94.5	0.33	0.26	7847	1.63	1.28
Student lives with you all of the time	BYP05	94.7	0.30	0.25	7814	1.39	1.18
Average number of people dependent on parent or spouse/partner	BYP06	2.5	0.02	0.01	7507	1.49	1.22
Married	BYP10	78.6	0.64	0.46	7819	1.90	1.38
Biological mother born in the U.S.	BYP17	95.0	0.38	0.25	7814	2.43	1.56
Parent's native language is English	BYP28	97.3	0.35	0.18	7765	3.67	1.92
Parent attended college, no 4-year degree	PARED	11.4	0.38	0.34	8757	1.24	1.11
Parent works full-time	BYP36	63.0	0.68	0.55	7802	1.55	1.24
Mother works in a professional occupation (groups a or b)	OCCUMOTH	19.9	0.63	0.43	8757	2.17	1.47
Spouse/partner works full-time	BYP40	85.3	0.51	0.44	6510	1.36	1.17
Father works in a professional occupation (groups a or b)	OCCUFATH	17.3	0.59	0.40	8757	2.09	1.45
Child changed school two times since first grade	BYP45	10.7	0.42	0.36	7443	1.39	1.18
Child was held back a grade	BYP46	9.9	0.45	0.35	7448	1.71	1.31
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	38.4	0.84	0.56	7420	2.23	1.49
Belong to school's parent-teacher organization	BYP54A	27.9	0.99	0.52	7371	3.56	1.89
Always discuss child's report card with child	BYP55B	88.2	0.44	0.37	7454	1.40	1.18
Provided advice or information about applying to college or other schools after high school	BYP56C	73.7	0.64	0.51	7432	1.57	1.25
Parent sometimes or frequently attended school activities	BYP57A	74.4	0.68	0.50	7472	1.80	1.34
Parent of child's friend did me a favor	BYP60B	71.7	0.66	0.52	7378	1.57	1.25
Child has another parent living outside of home	BYP61	30.5	0.67	0.53	7418	1.59	1.26
Average number of years living in current neighborhood	BYP65	11.4	0.15	0.10	7467	2.21	1.49
Low level of crime in neighborhood	BYP67	92.9	0.42	0.30	7419	1.98	1.41
Family rules enforced for doing homework	BYP69B	92.3	0.36	0.31	7405	1.35	1.16
Parent and child eat at least one meal together three times a week	BYP70	8.2	0.40	0.31	7775	1.66	1.29
Computer in home for 10th grader to use	BYP71	93.3	0.34	0.28	7807	1.48	1.22
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	52.9	1.61	0.90	3108	3.22	1.79
Child's school has a voice-messaging system	BYP75	53.0	1.34	0.58	7440	5.38	2.32
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.13	0.11	8757	1.39	1.18
Very important that child's school after high school is in a low crime environment	BYP80I	87.6	0.50	0.39	7140	1.67	1.29
Started a savings account for child's education after high school	BYP83A	76.0	0.84	0.68	3929	1.54	1.24
SUMMARY STATISTICS							
Mean							1.38
Minimum							1.11
Median					1.99		1.28
Maximum					1.24		2.32
Standard deviation					1.65		0.27
					5.38		
					0.90		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–9. Parent design effects – Multiracial

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	89.4	1.51	1.23	629	1.51	1.23
Student lives with you all of the time	BYP05	93.1	1.31	1.01	627	1.68	1.29
Average number of people dependent on parent or spouse/partner	BYP06	2.6	0.07	0.06	580	1.44	1.20
Married	BYP10	68.8	2.36	1.85	625	1.63	1.28
Biological mother born in the U.S.	BYP17	82.4	2.04	1.52	626	1.80	1.34
Parent's native language is English	BYP28	92.5	1.33	1.06	621	1.57	1.25
Parent attended college, no 4-year degree	PARED	15.2	1.43	1.32	742	1.18	1.09
Parent works full-time	BYP36	67.6	2.36	1.87	625	1.59	1.26
Mother works in a professional occupation (groups a or b)	OCCUMOTH	17.3	1.65	1.39	742	1.42	1.19
Spouse/partner works full-time	BYP40	81.0	2.46	1.80	477	1.86	1.37
Father works in a professional occupation (groups a or b)	OCCUFATH	16.6	1.58	1.37	742	1.34	1.16
Child changed school two times since first grade	BYP45	12.4	1.94	1.38	573	1.98	1.41
Child was held back a grade	BYP46	14.9	1.96	1.49	575	1.74	1.32
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	32.3	2.39	1.97	564	1.47	1.21
Belong to school's parent-teacher organization	BYP54A	24.3	2.38	1.80	569	1.74	1.32
Always discuss child's report card with child	BYP55B	87.5	1.56	1.39	569	1.27	1.12
Provided advice or information about applying to college or other schools after high school	BYP56C	72.3	2.80	1.88	566	2.21	1.49
Parent sometimes or frequently attended school activities	BYP57A	66.6	2.39	1.97	575	1.47	1.21
Parent of child's friend did me a favor	BYP60B	61.7	2.59	2.06	557	1.58	1.26
Child has another parent living outside of home	BYP61	45.0	2.75	2.10	564	1.73	1.31
Average number of years living in current neighborhood	BYP65	10.7	0.55	0.40	577	1.89	1.37
Low level of crime in neighborhood	BYP67	81.0	2.33	1.65	570	2.01	1.42
Family rules enforced for doing homework	BYP69B	93.6	1.21	1.03	569	1.37	1.17
Parent and child eat at least one meal together three times a week	BYP70	12.4	1.91	1.33	616	2.07	1.44
Computer in home for 10th grader to use	BYP71	86.4	1.70	1.38	620	1.52	1.23
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	57.1	4.34	3.34	221	1.69	1.30
Child's school has a voice-messaging system	BYP75	54.0	2.66	2.08	573	1.63	1.28
Want child to attend college, no 4-year degree	PARASPIR	0.9	0.36	0.34	742	1.09	1.05
Very important that child's school after high school is in a low crime environment	BYP80I	89.2	1.74	1.33	548	1.72	1.31
Started a savings account for child's education after high school	BYP83A	69.5	3.41	2.70	291	1.59	1.26
SUMMARY STATISTICS							
Mean							1.27
Minimum							1.05
Median					1.63		1.27
Maximum					1.09		1.49
Standard deviation					1.61		0.10
					2.21		
					0.26		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–10. Parent design effects – public

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	91.8	0.31	0.27	10518	1.31	1.14
Student lives with you all of the time	BYP05	94.6	0.26	0.22	10484	1.35	1.16
Average number of people dependent on parent or spouse/partner	BYP06	2.7	0.02	0.01	9692	1.96	1.40
Married	BYP10	71.6	0.55	0.44	10463	1.55	1.24
Biological mother born in the U.S.	BYP17	81.6	0.73	0.38	10434	3.68	1.92
Parent's native language is English	BYP28	84.5	0.73	0.35	10421	4.20	2.05
Parent attended college, no 4-year degree	PARED	12.1	0.34	0.30	12039	1.30	1.14
Parent works full-time	BYP36	63.6	0.56	0.47	10429	1.43	1.20
Mother works in a professional occupation (groups a or b)	OCCUMOTH	16.5	0.49	0.34	12039	2.08	1.44
Spouse/partner works full-time	BYP40	81.8	0.57	0.43	8001	1.75	1.32
Father works in a professional occupation (groups a or b)	OCCUFATH	13.5	0.46	0.31	12039	2.17	1.47
Child changed school two times since first grade	BYP45	12.0	0.38	0.33	9603	1.32	1.15
Child was held back a grade	BYP46	13.1	0.43	0.34	9644	1.56	1.25
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	32.3	0.68	0.48	9533	2.00	1.42
Belong to school's parent-teacher organization	BYP54A	22.5	0.79	0.43	9514	3.37	1.83
Always discuss child's report card with child	BYP55B	86.6	0.40	0.35	9616	1.31	1.15
Provided advice or information about applying to college or other schools after high school	BYP56C	73.2	0.54	0.45	9535	1.39	1.18
Parent sometimes or frequently attended school activities	BYP57A	68.2	0.60	0.47	9633	1.61	1.27
Parent of child's friend did me a favor	BYP60B	63.7	0.65	0.50	9420	1.71	1.31
Child has another parent living outside of home	BYP61	33.3	0.58	0.48	9528	1.45	1.20
Average number of years living in current neighborhood	BYP65	10.6	0.13	0.09	9621	1.96	1.40
Low level of crime in neighborhood	BYP67	88.4	0.51	0.33	9560	2.42	1.55
Family rules enforced for doing homework	BYP69B	92.9	0.30	0.26	9560	1.33	1.15
Parent and child eat at least one meal together three times a week	BYP70	8.9	0.34	0.28	10395	1.51	1.23
Computer in home for 10th grader to use	BYP71	86.0	0.52	0.34	10432	2.33	1.53
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	50.9	1.45	0.87	3278	2.77	1.67
Child's school has a voice-messaging system	BYP75	52.9	1.09	0.51	9609	4.55	2.13
Want child to attend college, no 4-year degree	PARASPIR	1.1	0.10	0.09	12039	1.13	1.06
Very important that child's school after high school is in a low crime environment	BYP80I	88.6	0.40	0.33	9059	1.43	1.20
Started a savings account for child's education after high school	BYP83A	76.3	0.76	0.65	4303	1.39	1.18
SUMMARY STATISTICS							
Mean						1.98	1.38
Minimum						1.13	1.06
Median						1.59	1.26
Maximum						4.55	2.13
Standard deviation						0.89	0.28

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–11. Parent design effects – Catholic

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	94.9	0.58	0.53	1746	1.21	1.10
Student lives with you all of the time	BYP05	96.3	0.52	0.45	1742	1.31	1.14
Average number of people dependent on parent or spouse/partner	BYP06	2.6	0.05	0.03	1661	2.28	1.51
Married	BYP10	84.5	0.93	0.87	1740	1.14	1.07
Biological mother born in the U.S.	BYP17	87.1	1.34	0.80	1737	2.79	1.67
Parent's native language is English	BYP28	91.7	1.32	0.66	1730	3.99	2.00
Parent attended college, no 4-year degree	PARED	12.0	0.82	0.74	1920	1.22	1.10
Parent works full-time	BYP36	59.2	1.53	1.18	1737	1.69	1.30
Mother works in a professional occupation (groups a or b)	OCCUMOTH	28.0	1.30	1.03	1920	1.60	1.26
Spouse/partner works full-time	BYP40	85.2	1.07	0.92	1476	1.34	1.16
Father works in a professional occupation (groups a or b)	OCCUFATH	26.1	1.45	1.00	1920	2.10	1.45
Child changed school two times since first grade	BYP45	11.8	0.88	0.79	1647	1.23	1.11
Child was held back a grade	BYP46	5.2	0.66	0.55	1643	1.46	1.21
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	65.0	1.72	1.18	1638	2.14	1.46
Belong to school's parent-teacher organization	BYP54A	40.5	2.03	1.22	1624	2.78	1.67
Always discuss child's report card with child	BYP55B	90.7	0.99	0.71	1645	1.93	1.39
Provided advice or information about applying to college or other schools after high school	BYP56C	81.5	1.36	0.96	1638	2.00	1.41
Parent sometimes or frequently attended school activities	BYP57A	82.7	1.21	0.93	1651	1.69	1.30
Parent of child's friend did me a favor	BYP60B	82.0	1.36	0.95	1625	2.03	1.43
Child has another parent living outside of home	BYP61	17.7	1.03	0.94	1652	1.20	1.10
Average number of years living in current neighborhood	BYP65	12.9	0.30	0.22	1650	1.88	1.37
Low level of crime in neighborhood	BYP67	89.2	0.99	0.77	1625	1.64	1.28
Family rules enforced for doing homework	BYP69B	95.1	0.63	0.53	1623	1.40	1.18
Parent and child eat at least one meal together three times a week	BYP70	7.4	0.71	0.63	1727	1.27	1.13
Computer in home for 10th grader to use	BYP71	98.0	0.45	0.34	1734	1.79	1.34
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	45.1	3.20	1.79	771	3.18	1.78
Child's school has a voice-messaging system	BYP75	72.4	2.75	1.10	1645	6.24	2.50
Want child to attend college, no 4-year degree	PARASPIR	0.2	0.09	0.09	1920	0.88	0.94
Very important that child's school after high school is in a low crime environment	BYP80I	87.1	0.74	0.84	1613	0.78	0.89
Started a savings account for child's education after high school	BYP83A	75.5	1.32	1.33	1045	0.99	0.99
SUMMARY STATISTICS							
Mean						1.91	1.34
Minimum						0.78	0.89
Median						1.67	1.29
Maximum						6.24	2.50
Standard deviation						1.08	0.33

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–12. Parent design effects – other private

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	95.1	0.64	0.62	1223	1.06	1.03
Student lives with you all of the time	BYP05	93.1	0.98	0.73	1214	1.81	1.34
Average number of people dependent on parent or spouse/partner	BYP06	2.5	0.07	0.04	1152	3.09	1.76
Married	BYP10	85.3	1.25	1.02	1219	1.52	1.23
Biological mother born in the U.S.	BYP17	84.6	2.44	1.03	1220	5.58	2.36
Parent's native language is English	BYP28	89.6	2.07	0.88	1210	5.59	2.37
Parent attended college, no 4-year degree	PARED	7.8	0.86	0.71	1403	1.44	1.20
Parent works full-time	BYP36	54.9	2.09	1.43	1212	2.15	1.46
Mother works in a professional occupation (groups a or b)	OCCUMOTH	25.7	1.94	1.17	1403	2.76	1.66
Spouse/partner works full-time	BYP40	83.4	1.28	1.16	1035	1.23	1.11
Father works in a professional occupation (groups a or b)	OCCUFATH	36.0	2.02	1.28	1403	2.49	1.58
Child changed school two times since first grade	BYP45	15.3	1.52	1.06	1144	2.03	1.42
Child was held back a grade	BYP46	9.1	0.83	0.85	1143	0.97	0.98
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	72.6	2.01	1.32	1137	2.31	1.52
Belong to school's parent-teacher organization	BYP54A	47.6	2.51	1.48	1133	2.85	1.69
Always discuss child's report card with child	BYP55B	84.9	1.58	1.06	1143	2.23	1.49
Provided advice or information about applying to college or other schools after high school	BYP56C	77.6	1.63	1.24	1137	1.74	1.32
Parent sometimes or frequently attended school activities	BYP57A	85.3	1.59	1.05	1144	2.30	1.52
Parent of child's friend did me a favor	BYP60B	83.1	1.26	1.12	1131	1.28	1.13
Child has another parent living outside of home	BYP61	18.0	1.79	1.14	1136	2.46	1.57
Average number of years living in current neighborhood	BYP65	10.5	0.45	0.23	1147	3.66	1.91
Low level of crime in neighborhood	BYP67	92.4	1.19	0.79	1138	2.28	1.51
Family rules enforced for doing homework	BYP69B	88.6	0.98	0.94	1137	1.08	1.04
Parent and child eat at least one meal together three times a week	BYP70	5.9	0.78	0.68	1194	1.30	1.14
Computer in home for 10th grader to use	BYP71	94.8	1.49	0.64	1210	5.42	2.33
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	60.9	4.23	2.09	546	4.09	2.02
Child's school has a voice-messaging system	BYP75	61.5	4.09	1.44	1138	8.03	2.83
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.43	0.27	1403	2.57	1.60
Very important that child's school after high school is in a low crime environment	BYP80I	87.8	1.37	0.98	1114	1.94	1.39
Started a savings account for child's education after high school	BYP83A	74.1	2.58	1.65	708	2.44	1.56
SUMMARY STATISTICS							
Mean						2.66	1.57
Minimum						0.97	0.98
Median						2.29	1.51
Maximum						8.03	2.83
Standard deviation						1.63	0.45

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–13. Parent design effects – low socioeconomic status (SES) quartile

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	89.1	0.66	0.56	3104	1.42	1.19
Student lives with you all of the time	BYP05	93.6	0.51	0.44	3096	1.36	1.17
Average number of people dependent on parent or spouse/partner	BYP06	2.9	0.04	0.03	2755	1.99	1.41
Married	BYP10	59.1	1.09	0.89	3082	1.52	1.23
Biological mother born in the U.S.	BYP17	66.0	1.76	0.86	3070	4.23	2.06
Parent's native language is English	BYP28	67.9	1.79	0.84	3073	4.52	2.13
Parent attended college, no 4-year degree	PARED	7.2	0.49	0.43	3635	1.33	1.16
Parent works full-time	BYP36	51.0	1.10	0.90	3066	1.47	1.21
Mother works in a professional occupation (groups a or b)	OCCUMOTH	1.6	0.24	0.21	3635	1.29	1.14
Spouse/partner works full-time	BYP40	69.4	1.33	1.01	2089	1.73	1.31
Father works in a professional occupation (groups a or b)	OCCUFATH	1.2	0.24	0.18	3635	1.69	1.30
Child changed school two times since first grade	BYP45	12.4	0.76	0.63	2712	1.44	1.20
Child was held back a grade	BYP46	21.8	0.97	0.79	2744	1.52	1.23
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	24.0	0.92	0.82	2685	1.24	1.12
Belong to school's parent-teacher organization	BYP54A	9.1	0.65	0.55	2701	1.39	1.18
Always discuss child's report card with child	BYP55B	81.8	0.86	0.74	2734	1.35	1.16
Provided advice or information about applying to college or other schools after high school	BYP56C	65.0	1.06	0.92	2678	1.31	1.14
Parent sometimes or frequently attended school activities	BYP57A	53.8	1.11	0.95	2728	1.34	1.16
Parent of child's friend did me a favor	BYP60B	51.8	1.20	0.97	2645	1.54	1.24
Child has another parent living outside of home	BYP61	36.3	1.15	0.93	2675	1.54	1.24
Average number of years living in current neighborhood	BYP65	10.4	0.24	0.19	2732	1.56	1.25
Low level of crime in neighborhood	BYP67	80.6	1.12	0.76	2704	2.15	1.47
Family rules enforced for doing homework	BYP69B	91.5	0.61	0.54	2701	1.28	1.13
Parent and child eat at least one meal together three times a week	BYP70	8.3	0.57	0.50	3053	1.28	1.13
Computer in home for 10th grader to use	BYP71	66.7	1.09	0.85	3069	1.65	1.29
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	36.5	2.72	2.28	447	1.42	1.19
Child's school has a voice-messaging system	BYP75	51.0	1.25	0.96	2728	1.71	1.31
Want child to attend college, no 4-year degree	PARASPIR	1.7	0.23	0.21	3635	1.14	1.07
Very important that child's school after high school is in a low crime environment	BYP80I	90.0	0.69	0.61	2407	1.26	1.12
Started a savings account for child's education after high school	BYP83A	72.2	2.28	1.75	659	1.71	1.31
SUMMARY STATISTICS							
Mean						1.68	1.27
Minimum						1.14	1.07
Median						1.46	1.21
Maximum						4.52	2.13
Standard deviation						0.77	0.24

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–14. Parent design effects – middle socioeconomic status (SES) quartiles

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	92.3	0.41	0.33	6399	1.54	1.24
Student lives with you all of the time	BYP05	95.0	0.36	0.27	6373	1.74	1.32
Average number of people dependent on parent or spouse/partner	BYP06	2.6	0.02	0.02	5949	1.49	1.22
Married	BYP10	71.6	0.66	0.57	6364	1.36	1.17
Biological mother born in the U.S.	BYP17	87.4	0.52	0.42	6359	1.56	1.25
Parent's native language is English	BYP28	90.3	0.50	0.37	6346	1.80	1.34
Parent attended college, no 4-year degree	PARED	18.4	0.51	0.45	7388	1.27	1.13
Parent works full-time	BYP36	68.5	0.69	0.58	6348	1.39	1.18
Mother works in a professional occupation (groups a or b)	OCCUMOTH	14.4	0.49	0.41	7388	1.46	1.21
Spouse/partner works full-time	BYP40	84.9	0.61	0.51	4895	1.41	1.19
Father works in a professional occupation (groups a or b)	OCCUFATH	8.6	0.38	0.33	7388	1.35	1.16
Child changed school two times since first grade	BYP45	11.9	0.48	0.42	5900	1.30	1.14
Child was held back a grade	BYP46	12.1	0.51	0.42	5917	1.47	1.21
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	32.4	0.81	0.61	5857	1.74	1.32
Belong to school's parent-teacher organization	BYP54A	21.2	0.73	0.53	5849	1.85	1.36
Always discuss child's report card with child	BYP55B	87.6	0.51	0.43	5908	1.40	1.18
Provided advice or information about applying to college or other schools after high school	BYP56C	73.4	0.66	0.58	5867	1.32	1.15
Parent sometimes or frequently attended school activities	BYP57A	70.7	0.79	0.59	5919	1.77	1.33
Parent of child's friend did me a favor	BYP60B	64.8	0.79	0.63	5800	1.60	1.27
Child has another parent living outside of home	BYP61	35.7	0.76	0.63	5868	1.48	1.22
Average number of years living in current neighborhood	BYP65	10.9	0.17	0.12	5914	2.05	1.43
Low level of crime in neighborhood	BYP67	88.7	0.57	0.41	5866	1.87	1.37
Family rules enforced for doing homework	BYP69B	93.8	0.35	0.31	5871	1.24	1.11
Parent and child eat at least one meal together three times a week	BYP70	9.0	0.47	0.36	6329	1.67	1.29
Computer in home for 10th grader to use	BYP71	90.5	0.45	0.37	6349	1.51	1.23
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	46.2	1.59	1.09	2103	2.14	1.46
Child's school has a voice-messaging system	BYP75	52.5	1.21	0.65	5892	3.46	1.86
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.14	0.12	7388	1.36	1.17
Very important that child's school after high school is in a low crime environment	BYP80I	90.1	0.45	0.40	5678	1.28	1.13
Started a savings account for child's education after high school	BYP83A	77.0	0.92	0.81	2718	1.30	1.14
SUMMARY STATISTICS							
Mean						1.61	1.26
Minimum						1.24	1.11
Median						1.48	1.22
Maximum						3.46	1.86
Standard deviation						0.42	0.15

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–15. Parent design effects – high socioeconomic status (SES) quartile

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	94.4	0.49	0.36	3984	1.83	1.35
Student lives with you all of the time	BYP05	94.8	0.47	0.35	3971	1.74	1.32
Average number of people dependent on parent or spouse/partner	BYP06	2.6	0.02	0.02	3801	1.58	1.26
Married	BYP10	87.2	0.70	0.53	3976	1.76	1.33
Biological mother born in the U.S.	BYP17	86.4	0.75	0.54	3962	1.89	1.37
Parent's native language is English	BYP28	91.0	0.68	0.46	3942	2.22	1.49
Parent attended college, no 4-year degree	PARED	3.7	0.37	0.29	4339	1.63	1.28
Parent works full-time	BYP36	64.1	1.14	0.76	3964	2.24	1.50
Mother works in a professional occupation (groups a or b)	OCCUMOTH	38.6	0.95	0.74	4339	1.63	1.28
Spouse/partner works full-time	BYP40	86.1	0.77	0.58	3528	1.76	1.33
Father works in a professional occupation (groups a or b)	OCCUFATH	40.6	1.00	0.75	4339	1.81	1.35
Child changed school two times since first grade	BYP45	12.2	0.72	0.53	3782	1.86	1.36
Child was held back a grade	BYP46	5.5	0.50	0.37	3769	1.81	1.34
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	50.0	1.21	0.81	3766	2.21	1.49
Belong to school's parent-teacher organization	BYP54A	43.2	1.45	0.81	3721	3.20	1.79
Always discuss child's report card with child	BYP55B	89.4	0.59	0.50	3762	1.36	1.17
Provided advice or information about applying to college or other schools after high school	BYP56C	82.1	0.82	0.63	3765	1.71	1.31
Parent sometimes or frequently attended school activities	BYP57A	80.8	0.83	0.64	3781	1.66	1.29
Parent of child's friend did me a favor	BYP60B	77.7	0.83	0.68	3731	1.48	1.21
Child has another parent living outside of home	BYP61	21.8	0.94	0.67	3773	1.94	1.39
Average number of years living in current neighborhood	BYP65	10.6	0.17	0.12	3772	1.95	1.39
Low level of crime in neighborhood	BYP67	95.4	0.49	0.34	3753	2.08	1.44
Family rules enforced for doing homework	BYP69B	92.3	0.55	0.44	3748	1.59	1.26
Parent and child eat at least one meal together three times a week	BYP70	8.6	0.52	0.45	3934	1.35	1.16
Computer in home for 10th grader to use	BYP71	98.6	0.25	0.19	3958	1.83	1.35
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	60.6	1.81	1.08	2045	2.79	1.67
Child's school has a voice-messaging system	BYP75	59.8	1.49	0.80	3772	3.48	1.87
Want child to attend college, no 4-year degree	PARASPIR	0.4	0.12	0.09	4339	1.66	1.29
Very important that child's school after high school is in a low crime environment	BYP80I	84.5	0.81	0.59	3701	1.84	1.35
Started a savings account for child's education after high school	BYP83A	76.3	1.03	0.82	2679	1.57	1.25
SUMMARY STATISTICS							
Mean						1.91	1.37
Minimum						1.35	1.16
Median						1.81	1.34
Maximum						3.48	1.87
Standard deviation						0.48	0.16

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–16. Parent design effects – urban

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	90.7	0.53	0.44	4446	1.47	1.21
Student lives with you all of the time	BYP05	94.5	0.48	0.34	4432	1.94	1.39
Average number of people dependent on parent or spouse/partner	BYP06	2.8	0.04	0.02	4027	3.45	1.86
Married	BYP10	67.0	1.08	0.71	4419	2.32	1.52
Biological mother born in the U.S.	BYP17	71.9	1.64	0.68	4411	5.88	2.43
Parent's native language is English	BYP28	76.1	1.64	0.64	4404	6.50	2.55
Parent attended college, no 4-year degree	PARED	12.6	0.65	0.46	5115	1.97	1.40
Parent works full-time	BYP36	62.3	0.94	0.73	4402	1.65	1.28
Mother works in a professional occupation (groups a or b)	OCCUMOTH	17.3	0.81	0.53	5115	2.36	1.54
Spouse/partner works full-time	BYP40	79.5	1.12	0.70	3292	2.54	1.59
Father works in a professional occupation (groups a or b)	OCCUFATH	15.7	0.73	0.51	5115	2.08	1.44
Child changed school two times since first grade	BYP45	14.0	0.77	0.55	3993	1.99	1.41
Child was held back a grade	BYP46	13.8	0.81	0.54	4007	2.19	1.48
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	36.7	1.22	0.77	3936	2.54	1.59
Belong to school's parent-teacher organization	BYP54A	25.8	1.24	0.70	3939	3.16	1.78
Always discuss child's report card with child	BYP55B	86.0	0.65	0.55	3982	1.38	1.17
Provided advice or information about applying to college or other schools after high school	BYP56C	75.4	1.00	0.69	3953	2.13	1.46
Parent sometimes or frequently attended school activities	BYP57A	65.8	1.04	0.75	3999	1.91	1.38
Parent of child's friend did me a favor	BYP60B	60.5	1.18	0.78	3893	2.27	1.51
Child has another parent living outside of home	BYP61	31.7	1.00	0.74	3966	1.85	1.36
Average number of years living in current neighborhood	BYP65	10.0	0.19	0.13	4001	2.17	1.47
Low level of crime in neighborhood	BYP67	81.3	1.15	0.62	3949	3.42	1.85
Family rules enforced for doing homework	BYP69B	92.8	0.54	0.41	3948	1.70	1.30
Parent and child eat at least one meal together three times a week	BYP70	8.9	0.53	0.43	4373	1.49	1.22
Computer in home for 10th grader to use	BYP71	83.5	0.97	0.56	4398	3.02	1.74
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	49.5	2.59	1.29	1498	4.02	2.00
Child's school has a voice-messaging system	BYP75	57.7	1.59	0.78	3989	4.14	2.03
Want child to attend college, no 4-year degree	PARASPIR	0.9	0.16	0.13	5115	1.46	1.21
Very important that child's school after high school is in a low crime environment	BYP80I	87.1	0.77	0.54	3819	2.03	1.42
Started a savings account for child's education after high school	BYP83A	75.9	1.34	0.95	2016	1.98	1.41
SUMMARY STATISTICS							
Mean						2.57	1.57
Minimum						1.38	1.17
Median						2.15	1.47
Maximum						6.50	2.55
Standard deviation						1.22	0.34

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–17. Parent design effects –suburban

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	93.0	0.41	0.32	6507	1.68	1.29
Student lives with you all of the time	BYP05	94.8	0.32	0.27	6490	1.35	1.16
Average number of people dependent on parent or spouse/partner	BYP06	2.6	0.02	0.02	6068	1.78	1.33
Married	BYP10	75.2	0.66	0.54	6477	1.52	1.23
Biological mother born in the U.S.	BYP17	83.4	0.84	0.46	6458	3.26	1.81
Parent's native language is English	BYP28	86.2	0.85	0.43	6447	3.89	1.97
Parent attended college, no 4-year degree	PARED	11.2	0.40	0.37	7399	1.20	1.10
Parent works full-time	BYP36	62.7	0.79	0.60	6465	1.73	1.32
Mother works in a professional occupation (groups a or b)	OCCUMOTH	17.8	0.68	0.44	7399	2.35	1.53
Spouse/partner works full-time	BYP40	82.9	0.73	0.52	5212	1.96	1.40
Father works in a professional occupation (groups a or b)	OCCUFATH	15.7	0.70	0.42	7399	2.75	1.66
Child changed school two times since first grade	BYP45	11.9	0.49	0.42	6012	1.35	1.16
Child was held back a grade	BYP46	11.6	0.53	0.41	6029	1.63	1.27
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	34.6	0.94	0.61	5993	2.34	1.53
Belong to school's parent-teacher organization	BYP54A	24.5	1.11	0.56	5964	3.97	1.99
Always discuss child's report card with child	BYP55B	87.1	0.53	0.43	6026	1.48	1.22
Provided advice or information about applying to college or other schools after high school	BYP56C	73.5	0.70	0.57	5990	1.53	1.24
Parent sometimes or frequently attended school activities	BYP57A	69.8	0.81	0.59	6032	1.89	1.37
Parent of child's friend did me a favor	BYP60B	66.9	0.89	0.61	5922	2.10	1.45
Child has another parent living outside of home	BYP61	30.8	0.75	0.60	5987	1.58	1.26
Average number of years living in current neighborhood	BYP65	10.7	0.15	0.11	6031	1.82	1.35
Low level of crime in neighborhood	BYP67	90.5	0.62	0.38	5992	2.64	1.62
Family rules enforced for doing homework	BYP69B	93.2	0.39	0.33	5994	1.46	1.21
Parent and child eat at least one meal together three times a week	BYP70	8.9	0.49	0.36	6430	1.88	1.37
Computer in home for 10th grader to use	BYP71	88.7	0.66	0.39	6457	2.77	1.67
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	53.5	1.88	1.04	2293	3.25	1.80
Child's school has a voice-messaging system	BYP75	55.9	1.34	0.64	6012	4.40	2.10
Want child to attend college, no 4-year degree	PARASPIR	1.0	0.13	0.12	7399	1.31	1.14
Very important that child's school after high school is in a low crime environment	BYP80I	88.4	0.53	0.42	5698	1.57	1.25
Started a savings account for child's education after high school	BYP83A	76.7	0.97	0.77	2989	1.58	1.26
SUMMARY STATISTICS							
Mean						2.13	1.44
Minimum						1.20	1.10
Median						1.80	1.34
Maximum						4.40	2.10
Standard deviation						0.86	0.27

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–18. Parent design effects –rural

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Biological parent completed questionnaire	BYP01	92.0	0.54	0.54	2534	1.00	1.00
Student lives with you all of the time	BYP05	94.1	0.56	0.47	2518	1.40	1.18
Average number of people dependent on parent or spouse/partner	BYP06	2.5	0.03	0.03	2410	1.25	1.12
Married	BYP10	74.5	1.17	0.87	2526	1.83	1.35
Biological mother born in the U.S.	BYP17	92.8	0.66	0.51	2522	1.65	1.28
Parent's native language is English	BYP28	95.1	0.55	0.43	2510	1.62	1.27
Parent attended college, no 4-year degree	PARED	12.6	0.72	0.62	2848	1.33	1.15
Parent works full-time	BYP36	65.1	1.04	0.95	2511	1.19	1.09
Mother works in a professional occupation (groups a or b)	OCCUMOTH	15.9	0.97	0.68	2848	2.02	1.42
Spouse/partner works full-time	BYP40	83.1	0.92	0.84	2008	1.22	1.11
Father works in a professional occupation (groups a or b)	OCCUFATH	10.9	0.75	0.58	2848	1.64	1.28
Child changed school two times since first grade	BYP45	9.9	0.62	0.61	2389	1.02	1.01
Child was held back a grade	BYP46	13.6	0.91	0.70	2394	1.68	1.30
Contacted by school about participating in school fundraising activities or doing volunteer work	BYP52H	34.6	1.30	0.98	2379	1.78	1.33
Belong to school's parent-teacher organization	BYP54A	21.3	1.62	0.84	2368	3.73	1.93
Always discuss child's report card with child	BYP55B	86.8	0.87	0.69	2396	1.60	1.26
Provided advice or information about applying to college or other schools after high school	BYP56C	72.1	0.98	0.92	2367	1.12	1.06
Parent sometimes or frequently attended school activities	BYP57A	73.5	1.24	0.90	2397	1.88	1.37
Parent of child's friend did me a favor	BYP60B	67.5	1.05	0.96	2361	1.19	1.09
Child has another parent living outside of home	BYP61	35.9	1.20	0.99	2363	1.47	1.21
Average number of years living in current neighborhood	BYP65	11.6	0.34	0.20	2386	2.94	1.71
Low level of crime in neighborhood	BYP67	93.9	0.69	0.49	2382	2.00	1.41
Family rules enforced for doing homework	BYP69B	92.2	0.61	0.55	2378	1.23	1.11
Parent and child eat at least one meal together three times a week	BYP70	8.2	0.66	0.55	2513	1.47	1.21
Computer in home for 10th grader to use	BYP71	87.2	0.98	0.67	2521	2.15	1.47
Use computer to communicate with 10th grader's teachers and administrative staff via E-mail about child	BYP74A	46.2	2.59	1.76	804	2.16	1.47
Child's school has a voice-messaging system	BYP75	44.7	3.03	1.02	2391	8.85	2.98
Want child to attend college, no 4-year degree	PARASPIR	1.2	0.21	0.20	2848	1.11	1.05
Very important that child's school after high school is in a low crime environment	BYP80I	90.7	0.60	0.61	2269	0.98	0.99
Started a savings account for child's education after high school	BYP83A	75.1	1.46	1.33	1051	1.20	1.09
SUMMARY STATISTICS							
Mean							1.31
Minimum							0.99
Median					1.86		1.24
Maximum					0.98		2.98
Standard deviation					1.53		0.38
					8.85		
					1.45		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–19. School design effects – all

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
School enrollment is between 1,000 - 1,199 students	BYSCENP	5.5	1.04	0.91	630	1.31	1.15
Grade spans: 6, 7, or 8 through 12 or higher; and 9 through 10, 11, 12, or higher	BYSPANP	68.1	3.38	1.71	743	3.91	1.98
Year round school	BYA03E	4.4	1.48	0.76	737	3.82	1.96
Offer differentiated courses in core curriculum but students have open access to any course	BYA04	53.1	3.39	1.97	643	2.96	1.72
Semester system	BYA06	70.0	3.07	1.79	653	2.93	1.71
Average number of days in school year for 10th graders	BYA07	179.4	0.19	0.14	735	1.97	1.41
Average number of class periods in school day for 10th graders	BYA08	6.3	0.08	0.06	736	2.27	1.51
Average number of minutes of class period for 10th graders	BYA09	59.2	1.01	0.67	738	2.30	1.52
Average size of full academic class load for 10th graders	BYA10	6.2	0.09	0.05	652	3.01	1.74
School is co-educational	BYA11	97.3	0.53	0.60	740	0.77	0.88
Average percentage of students participating in alcohol/drug prevention program	BYA12G	50.8	2.86	1.67	524	2.93	1.71
Parents not notified when students are absent without excuse	BYA13	5.8	1.41	0.92	647	2.35	1.53
Average number of 10th graders in a special education program	BYA14E	8.8	0.69	0.37	698	3.41	1.85
No vocational-technical programs or services offered	BYA16	25.7	2.73	1.72	648	2.52	1.59
Basketball offered for male or female students	BYA19CA AND BYA19CB	96.5	1.46	0.74	611	3.88	1.97
Average percentage of 10th graders with limited English proficiency	BYA20	2.3	0.34	0.24	713	2.02	1.42
Percent of 10th graders receiving free or reduced-price lunch: 21-30	BY10FLP	14.2	2.32	1.34	685	3.03	1.74
Number of full-time teachers: 31-45	BYFTTP	15.0	2.23	1.34	713	2.78	1.67
Average number of part-time teachers	BYA22B	4.0	0.38	0.19	687	3.93	1.98
Average number of full-time math teachers	BYA23A	5.0	0.21	0.18	632	1.37	1.17
Average percentage of full-time teachers who are certified	BYA24A	87.7	1.78	0.97	721	3.38	1.84
Average percentage of certified full-time teachers teaching outside their field	BYA25A	4.9	0.97	0.60	603	2.58	1.61
Average lowest annual salary for full-time teachers	BYA26A	25711.6	417.13	252.93	590	2.72	1.65
Good teachers are given priority on requests for materials	BYA28E	3.4	1.20	0.71	643	2.83	1.68
Content standards for academic subjects linked to performance standards for assessment of students' mastery of content	BYA31	82.6	2.43	1.53	615	2.52	1.59
Students not required to pass minimum competency or proficiency test to receive high school diploma	BYA32	49.6	3.07	1.96	652	2.45	1.57
Performance of students on standardized tests has great deal of influence on how principal's performance is evaluated	BYA48A	50.3	3.39	2.02	614	2.82	1.68
Learning of 10th graders hindered not at all by lack of discipline and safety	BYA50J	58.5	3.54	1.99	616	3.17	1.78
Library media center is centralized	BYL01	92.6	1.96	0.98	708	3.96	1.99
Parents not allowed to check out materials from the library media center	BYL28	27.8	2.92	1.70	698	2.95	1.72
SUMMARY STATISTICS							
Mean							1.64
Minimum							0.88
Median					2.76		1.68
Maximum					0.77		1.99
Standard deviation					2.83		0.26
					3.96		
					0.78		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–20. School design effects –public

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
School enrollment is between 1,000 - 1,199 students	BYSCENP	5.6	0.85	1.05	483	0.65	0.81
Grade spans: 6, 7, or 8 through 12 or higher; and 9 through 10, 11, 12, or higher	BYSPANP	80.2	4.21	1.67	572	6.37	2.52
Year round school	BYA03E	2.8	1.11	0.69	569	2.55	1.60
Offer differentiated courses in core curriculum but students have open access to any course	BYA04	62.4	3.96	2.18	494	3.29	1.81
Semester system	BYA06	77.5	3.25	1.87	502	3.04	1.74
Average number of days in school year for 10th graders	BYA07	179.5	0.21	0.14	566	2.21	1.49
Average number of class periods in school day for 10th graders	BYA08	6.2	0.10	0.07	567	2.38	1.54
Average number of minutes of class period for 10th graders	BYA09	62.0	1.31	0.81	568	2.61	1.62
Average size of full academic class load for 10th graders	BYA10	6.2	0.10	0.06	502	3.07	1.75
School is co-educational	BYA11	99.9	0.11	0.14	569	0.65	0.80
Average percentage of students participating in alcohol/drug prevention program	BYA12G	46.7	3.26	1.82	412	3.20	1.79
Parents not notified when students are absent without excuse	BYA13	5.7	1.45	1.04	494	1.92	1.39
Average number of 10th graders in a special education program	BYA14E	11.3	0.86	0.41	539	4.30	2.07
No vocational-technical programs or services offered	BYA16	11.8	2.62	1.44	500	3.30	1.82
Basketball offered for male or female students	BYA19CA and BYA19CB	98.4	0.93	0.56	496	2.75	1.66
Average percentage of 10th graders with limited English proficiency	BYA20	2.4	0.32	0.27	545	1.33	1.15
Percent of 10th graders receiving free or reduced-price lunch: 21-30	BY10FLP	19.5	3.12	1.74	517	3.20	1.79
Number of full-time teachers: 31-45	BYFTTP	15.9	2.69	1.56	550	2.98	1.73
Average number of part-time teachers	BYA22B	3.8	0.49	0.23	524	4.37	2.09
Average number of full-time math teachers	BYA23A	5.7	0.28	0.22	483	1.58	1.26
Average percentage of full-time teachers who are certified	BYA24A	96.6	1.11	0.46	562	5.93	2.44
Average percentage of certified full-time teachers teaching outside their field	BYA25A	4.8	1.16	0.70	469	2.73	1.65
Average lowest annual salary for full-time teachers	BYA26A	27481.8	284.48	205.26	457	1.92	1.39
Good teachers are given priority on requests for materials	BYA28E	1.8	0.67	0.60	492	1.26	1.12
Content standards for academic subjects linked to performance standards for assessment of students' mastery of content	BYA31	88.9	2.48	1.43	482	2.98	1.73
Students not required to pass minimum competency or proficiency test to receive high school diploma	BYA32	43.6	3.31	2.22	500	2.23	1.49
Performance of students on standardized tests has great deal of influence on how principal's performance is evaluated	BYA48A	56.8	3.85	2.28	473	2.84	1.69
Learning of 10th graders hindered not at all by lack of discipline and safety	BYA50J	52.7	4.12	2.30	474	3.23	1.80
Library media center is centralized	BYL01	96.4	1.57	0.79	554	3.91	1.98
Parents not allowed to check out materials from the library media center	BYL28	27.6	3.35	1.91	547	3.06	1.75
SUMMARY STATISTICS							
Mean						2.86	1.65
Minimum						0.65	0.80
Median						2.91	1.71
Maximum						6.37	2.52
Standard deviation						1.28	0.39

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–21. School design effects – private

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
School enrollment is between 1,000 - 1,199 students	BYSCENP	5.1	3.34	1.83	147	3.34	1.83
Grade spans: 6, 7, or 8 through 12 or higher; and 9 through 10, 11, 12, or higher	BYSPANP	33.5	4.34	3.62	171	1.44	1.20
Year round school	BYA03E	9.0	4.69	2.21	168	4.50	2.12
Offer differentiated courses in core curriculum but students have open access to any course	BYA04	24.9	5.48	3.55	149	2.38	1.54
Semester system	BYA06	47.9	6.95	4.08	151	2.91	1.70
Average number of days in school year for 10th graders	BYA07	179.0	0.44	0.34	169	1.62	1.27
Average number of class periods in school day for 10th graders	BYA08	6.7	0.15	0.09	169	2.75	1.66
Average number of minutes of class period for 10th graders	BYA09	50.8	0.88	0.83	170	1.11	1.05
Average size of full academic class load for 10th graders	BYA10	6.2	0.15	0.09	150	2.91	1.71
School is co-educational	BYA11	89.8	2.19	2.33	171	0.88	0.94
Average percentage of students participating in alcohol/drug prevention program	BYA12G	69.5	5.29	3.64	112	2.10	1.45
Parents not notified when students are absent without excuse	BYA13	6.0	3.51	1.93	153	3.32	1.82
Average number of 10th graders in a special education program	BYA14E	1.3	0.54	0.49	159	1.23	1.11
No vocational-technical programs or services offered	BYA16	69.0	7.00	3.81	148	3.37	1.83
Basketball offered for male or female students	BYA19CA and BYA19CB	90.3	5.31	2.78	115	3.66	1.91
Average percentage of 10th graders with limited English proficiency	BYA20	2.0	0.97	0.51	168	3.60	1.90
Percent of 10th graders receiving free or reduced-price lunch: 21-30	BY10FLP	0.5	0.40	0.57	168	0.49	0.70
Number of full-time teachers: 31-45	BYFTTP	12.5	3.77	2.60	163	2.10	1.45
Average number of part-time teachers	BYA22B	4.7	0.45	0.29	163	2.41	1.55
Average number of full-time math teachers	BYA23A	2.7	0.28	0.22	149	1.63	1.28
Average percentage of full-time teachers who are certified	BYA24A	59.9	5.45	2.98	159	3.33	1.82
Average percentage of certified full-time teachers teaching outside their field	BYA25A	5.5	1.60	1.14	134	1.94	1.39
Average lowest annual salary for full-time teachers	BYA26A	19767.2	1107.62	638.64	133	3.01	1.73
Good teachers are given priority on requests for materials	BYA28E	8.1	4.21	2.23	151	3.58	1.89
Content standards for academic subjects linked to performance standards for assessment of students' mastery of content	BYA31	63.5	6.66	4.19	133	2.53	1.59
Students not required to pass minimum competency or proficiency test to receive high school diploma	BYA32	67.2	7.37	3.82	152	3.72	1.93
Performance of students on standardized tests has great deal of influence on how principal's performance is evaluated	BYA48A	31.9	7.21	3.94	141	3.35	1.83
Learning of 10th graders hindered not at all by lack of discipline and safety	BYA50J	75.1	6.48	3.64	142	3.16	1.78
Library media center is centralized	BYL01	79.4	6.47	3.27	154	3.92	1.98
Parents not allowed to check out materials from the library media center	BYL28	28.6	5.90	3.69	151	2.56	1.60
SUMMARY STATISTICS							
Mean						2.63	1.59
Minimum						0.49	0.70
Median						2.83	1.68
Maximum						4.50	2.12
Standard deviation						1.00	0.34

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–22. School design effects – small 10th

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
School enrollment is between 1,000 - 1,199 students	BYSCENP	4.8	1.18	1.13	356	1.10	1.05
Grade spans: 6, 7, or 8 through 12 or higher; and 9 through 10, 11, 12, or higher	BYSPANP	63.0	3.90	2.40	406	2.65	1.63
Year round school	BYA03E	4.5	1.76	1.04	401	2.89	1.70
Offer differentiated courses in core curriculum but students have open access to any course	BYA04	50.2	3.90	2.63	363	2.21	1.49
Semester system	BYA06	67.7	3.58	2.44	367	2.15	1.47
Average number of days in school year for 10th graders	BYA07	179.3	0.23	0.18	402	1.53	1.24
Average number of class periods in school day for 10th graders	BYA08	6.4	0.10	0.07	403	1.76	1.33
Average number of minutes of class period for 10th graders	BYA09	58.3	1.21	0.89	403	1.83	1.35
Average size of full academic class load for 10th graders	BYA10	6.2	0.10	0.07	366	2.32	1.52
School is co-educational	BYA11	96.9	0.64	0.87	406	0.54	0.74
Average percentage of students participating in alcohol/drug prevention program	BYA12G	53.6	3.33	2.26	287	2.17	1.47
Parents not notified when students are absent without excuse	BYA13	5.8	1.64	1.22	365	1.79	1.34
Average number of 10th graders in a special education program	BYA14E	8.5	0.82	0.52	383	2.47	1.57
No vocational-technical programs or services offered	BYA16	29.4	3.26	2.39	366	1.87	1.37
Basketball offered for male or female students	BYA19CA AND BYA19CB	95.8	1.74	1.10	331	2.52	1.59
Average percentage of 10th graders with limited English proficiency	BYA20	1.7	0.40	0.30	396	1.74	1.32
Percent of 10th graders receiving free or reduced-price lunch: 21-30	BY10FLP	14.0	2.75	1.78	380	2.38	1.54
Number of full-time teachers: 31-45	BYFTTP	17.9	2.71	1.95	386	1.93	1.39
Average number of part-time teachers	BYA22B	4.0	0.45	0.26	376	2.99	1.73
Average number of full-time math teachers	BYA23A	3.5	0.17	0.14	360	1.43	1.20
Average percentage of full-time teachers who are certified	BYA24A	86.1	2.13	1.39	395	2.34	1.53
Average percentage of certified full-time teachers teaching outside their field	BYA25A	5.3	1.13	0.82	342	1.90	1.38
Average lowest annual salary for full-time teachers	BYA26A	24925.3	479.90	332.41	332	2.08	1.44
Good teachers are given priority on requests for materials	BYA28E	3.8	1.42	1.01	360	1.97	1.41
Content standards for academic subjects linked to performance standards for assessment of students' mastery of content	BYA31	80.8	2.90	2.14	339	1.84	1.36
Students not required to pass minimum competency or proficiency test to receive high school diploma	BYA32	53.5	3.65	2.61	367	1.96	1.40
Performance of students on standardized tests has great deal of influence on how principal's performance is evaluated	BYA48A	46.8	3.98	2.69	344	2.18	1.48
Learning of 10th graders hindered not at all by lack of discipline and safety	BYA50J	60.4	4.20	2.64	345	2.54	1.59
Library media center is centralized	BYL01	91.5	2.35	1.42	385	2.73	1.65
Parents not allowed to check out materials from the library media center	BYL28	25.7	3.44	2.25	379	2.34	1.53
SUMMARY STATISTICS							
Mean						2.07	1.43
Minimum						0.54	0.74
Median						2.12	1.45
Maximum						2.99	1.73
Standard deviation						0.52	0.20

NOTE: Small schools are defined as those with 10th-grade enrollment less than 300.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–23. School design effects – large 10th

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
School enrollment is between 1,000 - 1,199 students	BYSCENP	9.5	2.12	1.78	274	1.42	1.19
Grade spans: 6, 7, or 8 through 12 or higher; and 9 through 10, 11, 12, or higher	BYSPANP	93.7	1.08	1.32	337	0.66	0.82
Year round school	BYA03E	3.9	1.18	1.05	336	1.25	1.12
Offer differentiated courses in core curriculum but students have open access to any course	BYA04	69.0	3.37	2.77	280	1.48	1.21
Semester system	BYA06	82.3	2.40	2.26	286	1.12	1.06
Average number of days in school year for 10th graders	BYA07	179.5	0.19	0.20	333	0.94	0.97
Average number of class periods in school day for 10th graders	BYA08	5.9	0.08	0.08	333	0.92	0.96
Average number of minutes of class period for 10th graders	BYA09	63.8	1.00	1.03	335	0.94	0.97
Average size of full academic class load for 10th graders	BYA10	6.0	0.09	0.08	286	1.34	1.16
School is co-educational	BYA11	99.4	0.26	0.43	334	0.36	0.60
Average percentage of students participating in alcohol/drug prevention program	BYA12G	37.9	2.40	2.31	237	1.08	1.04
Parents not notified when students are absent without excuse	BYA13	5.7	1.55	1.38	282	1.25	1.12
Average number of 10th graders in a special education program	BYA14E	10.7	0.43	0.42	315	1.04	1.02
No vocational-technical programs or services offered	BYA16	5.4	1.31	1.35	282	0.94	0.97
Basketball offered for male or female students	BYA19CA and BYA19CB	100.0	0.00	0.00	280	#	#
Average percentage of 10th graders with limited English proficiency	BYA20	5.5	0.42	0.45	317	0.88	0.94
Percent of 10th graders receiving free or reduced-price lunch: 21-30	BY10FLP	15.5	2.33	2.08	305	1.26	1.12
Number of full-time teachers: 31-45	BYFTTP	1.2	0.60	0.61	327	0.97	0.99
Average number of part-time teachers	BYA22B	4.2	0.27	0.27	311	1.02	1.01
Average number of full-time math teachers	BYA23A	12.9	0.23	0.27	272	0.70	0.83
Average percentage of full-time teachers who are certified	BYA24A	96.2	0.61	0.68	326	0.80	0.90
Average percentage of certified full-time teachers teaching outside their field	BYA25A	2.9	0.80	0.77	261	1.08	1.04
Average lowest annual salary for full-time teachers	BYA26A	29877.9	312.54	298.54	258	1.10	1.05
Good teachers are given priority on requests for materials	BYA28E	1.0	0.59	0.60	283	0.97	0.99
Content standards for academic subjects linked to performance standards for assessment of students' mastery of content	BYA31	91.9	1.87	1.65	276	1.29	1.14
Students not required to pass minimum competency or proficiency test to receive high school diploma	BYA32	28.3	2.59	2.67	285	0.94	0.97
Performance of students on standardized tests has great deal of influence on how principal's performance is evaluated	BYA48A	69.4	3.03	2.81	270	1.16	1.08
Learning of 10th graders hindered not at all by lack of discipline and safety	BYA50J	48.2	3.48	3.04	271	1.31	1.14
Library media center is centralized	BYL01	98.0	0.73	0.78	323	0.87	0.93
Parents not allowed to check out materials from the library media center	BYL28	37.9	2.97	2.72	319	1.19	1.09
SUMMARY STATISTICS							
Mean						1.04	1.01
Minimum						0.36	0.60
Median						1.04	1.02
Maximum						1.48	1.21
Standard deviation						0.24	0.13

#The design effect is undefined because the estimate is 100.00.
NOTE: Large schools are defined as those with 10th-grade enrollment of at least 300.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–24. Student design effects – all

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	34.3	0.66	0.38	15362	3.01	1.73
Student born in 1986	DOBIRTHP	57.4	0.54	0.40	15308	1.83	1.35
Student lives with father and mother	BYFCOMP	56.8	0.57	0.40	15362	2.01	1.42
Reading test number-right score	BYTXRIRR	29.5	0.18	0.08	15362	5.18	2.28
Mathematics test number-right score	BYTXMIRR	37.2	0.23	0.10	15362	5.60	2.37
Composite achievement test highest quartile	BYTXCQU	25.0	0.68	0.35	15362	3.77	1.94
Agree or strongly agree that school rules are fair	BYS21B	54.1	0.65	0.41	14494	2.47	1.57
Never was hit the first semester	BYS22E	78.6	0.46	0.34	14590	1.80	1.34
Someone bullied or picked on student at least once	BYS22H	20.2	0.44	0.33	14654	1.74	1.32
High school program is general	SCHPROG	38.6	0.63	0.39	15362	2.56	1.60
Average number of hours spent on homework each week out of school	BYS34B	5.7	0.08	0.05	14903	2.95	1.72
Average number of hours spent on English homework each week out of school	BYS36B	2.4	0.04	0.03	13913	2.30	1.52
Good grades not important	BYS37	1.4	0.12	0.09	15086	1.50	1.22
Never come to class without books	BYS38B	51.6	0.62	0.42	14482	2.21	1.49
Participated in band, orchestra, chorus, or choir	BYS41A	21.5	0.52	0.34	15011	2.38	1.54
Did not participate in a hobby club	BYS41H	90.5	0.34	0.24	14942	2.04	1.43
Use computer as a resource to learn things of interest at least once a week	BYS45C	42.4	0.57	0.41	14317	1.87	1.37
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	52.5	0.57	0.42	14163	1.85	1.36
Being able to give own children better opportunities very important	BYS54G	80.3	0.43	0.33	14637	1.68	1.30
Expect to finish college, but not advanced degree	STEXPECT	35.8	0.46	0.39	15362	1.42	1.19
Plan to continue education right after high school	BYS57	71.6	0.52	0.38	13972	1.87	1.37
Hope to receive an athletic scholarship	BYS61	73.2	0.74	0.57	6142	1.70	1.31
At age 30 exp to be a manager	BYOCC30	1.9	0.14	0.12	13520	1.48	1.22
At age 30 exp to be in the military	BYOCC30	0.9	0.10	0.08	13520	1.53	1.24
At age 30 exp to be an operative	BYOCC30	0.6	0.10	0.06	13520	2.42	1.55
At age 30 exp to be a professional (group b)	BYOCC30	20.4	0.43	0.35	13520	1.54	1.24
At age 30 exp to be a technician	BYOCC30	3.3	0.19	0.15	13520	1.59	1.26
At age 30 doesn't know what to be	BYOCC30	34.6	0.53	0.41	13520	1.68	1.30
English is native language	STLANG	86.0	0.60	0.28	15362	4.61	2.15
Among close friends, somewhat or very important that they have a regular job	BYS90K	79.0	0.56	0.40	10472	1.98	1.41
SUMMARY STATISTICS							
Mean							1.50
Minimum							1.19
Median							1.39
Maximum					2.35		2.37
Standard deviation					1.42		0.31
					1.93		
					5.60		
					1.09		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–25. Student design effects – male

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	33.8	0.83	0.54	7646	2.38	1.54
Student born in 1986	DOBIRTHP	53.0	0.76	0.57	7606	1.75	1.32
Student lives with father and mother	BYFCOMP	57.1	0.72	0.57	7646	1.62	1.27
Reading test number-right score	BYTXRIRR	28.9	0.20	0.12	7646	3.03	1.74
Mathematics test number-right score	BYTXMIRR	37.8	0.26	0.14	7646	3.33	1.82
Composite achievement test highest quartile	BYTXCQU	25.6	0.78	0.50	7646	2.45	1.57
Agree or strongly agree that school rules are fair	BYS21B	51.9	0.82	0.59	7186	1.92	1.39
Never was hit the first semester	BYS22E	70.6	0.71	0.54	7228	1.77	1.33
Someone bullied or picked on student at least once	BYS22H	19.7	0.61	0.47	7266	1.70	1.30
High school program is general	SCHPROG	39.3	0.79	0.56	7646	2.00	1.41
Average number of hours spent on homework each week out of school	BYS34B	5.1	0.09	0.06	7353	2.18	1.48
Average number of hours spent on English homework each week out of school	BYS36B	2.2	0.05	0.04	6857	2.04	1.43
Good grades not important	BYS37	2.1	0.20	0.16	7477	1.53	1.24
Never come to class without books	BYS38B	50.0	0.73	0.59	7153	1.51	1.23
Participated in band, orchestra, chorus, or choir	BYS41A	16.3	0.60	0.43	7430	1.95	1.40
Did not participate in a hobby club	BYS41H	91.9	0.41	0.32	7396	1.65	1.29
Use computer as a resource to learn things of interest at least once a week	BYS45C	45.3	0.78	0.59	7056	1.75	1.32
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	54.2	0.79	0.60	6979	1.74	1.32
Being able to give own children better opportunities very important	BYS54G	78.9	0.63	0.48	7195	1.74	1.32
Expect to finish college, but not advanced degree	STEXPECT	37.2	0.66	0.55	7646	1.44	1.20
Plan to continue education right after high school	BYS57	66.4	0.73	0.58	6713	1.62	1.27
Hope to receive an athletic scholarship	BYS61	77.9	0.94	0.72	3335	1.72	1.31
At age 30 exp to be a manager	BYOCC30	2.2	0.22	0.18	6529	1.51	1.23
At age 30 exp to be in the military	BYOCC30	1.7	0.19	0.16	6529	1.51	1.23
At age 30 exp to be an operative	BYOCC30	1.1	0.19	0.13	6529	2.29	1.51
At age 30 exp to be a professional (group b)	BYOCC30	11.8	0.50	0.40	6529	1.54	1.24
At age 30 exp to be a technician	BYOCC30	4.5	0.32	0.26	6529	1.55	1.24
At age 30 doesn't know what to be	BYOCC30	38.9	0.73	0.60	6529	1.48	1.22
English is native language	STLANG	86.2	0.62	0.39	7646	2.49	1.58
Among close friends, somewhat or very important that they have a regular job	BYS90K	81.8	0.70	0.55	4968	1.65	1.29
SUMMARY STATISTICS							
Mean							1.37
Minimum							1.20
Median							1.32
Maximum					1.90		1.82
Standard deviation					1.44		0.16
					1.74		
					3.33		
					0.46		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–26. Student design effects – female

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	34.8	0.81	0.54	7716	2.23	1.49
Student born in 1986	DOBIRTHP	62.0	0.71	0.55	7702	1.64	1.28
Student lives with father and mother	BYFCOMP	56.5	0.79	0.56	7716	1.96	1.40
Reading test number-right score	BYTXRIRR	30.1	0.21	0.11	7716	3.88	1.97
Mathematics test number-right score	BYTXMIRR	36.6	0.26	0.13	7716	3.88	1.97
Composite achievement test highest quartile	BYTXCQU	24.5	0.84	0.49	7716	2.97	1.72
Agree or strongly agree that school rules are fair	BYS21B	56.2	0.92	0.58	7308	2.49	1.58
Never was hit the first semester	BYS22E	86.7	0.51	0.40	7362	1.68	1.30
Someone bullied or picked on student at least once	BYS22H	20.6	0.60	0.47	7388	1.60	1.27
High school program is general	SCHPROG	37.8	0.79	0.55	7716	2.06	1.44
Average number of hours spent on homework each week out of school	BYS34B	6.3	0.10	0.07	7550	2.28	1.51
Average number of hours spent on English homework each week out of school	BYS36B	2.7	0.05	0.04	7056	1.96	1.40
Good grades not important	BYS37	0.7	0.10	0.09	7609	1.20	1.10
Never come to class without books	BYS38B	53.2	0.86	0.58	7329	2.18	1.48
Participated in band, orchestra, chorus, or choir	BYS41A	26.8	0.71	0.51	7581	1.94	1.39
Did not participate in a hobby club	BYS41H	89.1	0.50	0.36	7546	1.95	1.40
Use computer as a resource to learn things of interest at least once a week	BYS45C	39.4	0.71	0.57	7261	1.55	1.24
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	50.9	0.76	0.59	7184	1.65	1.28
Being able to give own children better opportunities very important	BYS54G	81.6	0.57	0.45	7442	1.62	1.27
Expect to finish college, but not advanced degree	STEXPECT	34.4	0.65	0.54	7716	1.44	1.20
Plan to continue education right after high school	BYS57	76.5	0.67	0.50	7259	1.79	1.34
Hope to receive an athletic scholarship	BYS61	67.6	1.03	0.88	2807	1.36	1.17
At age 30 exp to be a manager	BYOCC30	1.7	0.19	0.15	6991	1.52	1.23
At age 30 exp to be in the military	BYOCC30	0.2	0.08	0.06	6991	1.79	1.34
At age 30 exp to be an operative	BYOCC30	0.1	0.04	0.03	6991	1.54	1.24
At age 30 exp to be a professional (group b)	BYOCC30	28.7	0.61	0.54	6991	1.29	1.14
At age 30 exp to be a technician	BYOCC30	2.2	0.23	0.17	6991	1.67	1.29
At age 30 doesn't know what to be	BYOCC30	30.5	0.68	0.55	6991	1.54	1.24
English is native language	STLANG	85.7	0.81	0.40	7716	4.12	2.03
Among close friends, somewhat or very important that they have a regular job	BYS90K	76.5	0.70	0.57	5504	1.52	1.23
SUMMARY STATISTICS							
Mean							1.40
Minimum							1.10
Median							1.32
Maximum					2.01		2.03
Standard deviation					1.20		0.24
					1.73		
					4.12		
					0.76		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–27. Student design effects – Indian

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	18.9	5.69	3.43	131	2.75	1.66
Student born in 1986	DOBIRTHP	44.8	4.38	4.36	131	1.01	1.00
Student lives with father and mother	BYFCOMP	43.5	5.41	4.35	131	1.55	1.24
Reading test number-right score	BYTXRIRR	26.0	0.80	0.73	131	1.22	1.10
Mathematics test number-right score	BYTXMIRR	32.2	1.08	0.81	131	1.81	1.35
Composite achievement test highest quartile	BYTXCQU	4.1	1.73	1.73	131	1.00	1.00
Agree or strongly agree that school rules are fair	BYS21B	48.3	4.81	4.52	123	1.13	1.06
Never was hit the first semester	BYS22E	70.0	4.83	4.13	124	1.37	1.17
Someone bullied or picked on student at least once	BYS22H	19.5	4.45	3.57	124	1.55	1.24
High school program is general	SCHPROG	44.6	5.18	4.36	131	1.41	1.19
Average number of hours spent on homework each week out of school	BYS34B	5.8	0.79	0.61	125	1.70	1.30
Average number of hours spent on English homework each week out of school	BYS36B	2.4	0.31	0.30	119	1.09	1.04
Good grades not important	BYS37	6.2	3.24	2.14	127	2.29	1.51
Never come to class without books	BYS38B	46.3	5.84	4.55	121	1.65	1.28
Participated in band, orchestra, chorus, or choir	BYS41A	12.3	3.75	2.93	127	1.63	1.28
Did not participate in a hobby club	BYS41H	94.7	2.23	2.00	126	1.24	1.12
Use computer as a resource to learn things of interest at least once a week	BYS45C	37.3	4.47	4.47	118	1.00	1.00
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	54.7	5.30	4.68	114	1.28	1.13
Being able to give own children better opportunities very important	BYS54G	87.8	3.54	3.02	119	1.38	1.17
Expect to finish college, but not advanced degree	STEXPECT	30.9	6.15	4.05	131	2.30	1.52
Plan to continue education right after high school	BYS57	56.6	4.58	4.66	114	0.97	0.98
Hope to receive an athletic scholarship	BYS61	76.3	5.14	5.36	64	0.92	0.96
At age 30 exp to be a manager	BYOCC30	0.0	0.00	0.00	112	#	#
At age 30 exp to be in the military	BYOCC30	0.8	0.79	0.83	112	0.90	0.95
At age 30 exp to be an operative	BYOCC30	0.0	0.00	0.00	112	#	#
At age 30 exp to be a professional (group b)	BYOCC30	13.4	3.42	3.24	112	1.12	1.06
At age 30 exp to be a technician	BYOCC30	5.7	2.40	2.19	112	1.20	1.10
At age 30 doesn't know what to be	BYOCC30	43.6	4.88	4.71	112	1.07	1.04
English is native language	STLANG	83.7	4.46	3.24	131	1.90	1.38
Among close friends, somewhat or very important that they have a regular job	BYS90K	76.6	6.06	5.06	71	1.44	1.20
SUMMARY STATISTICS							
Mean							1.18
Minimum							0.95
Median							1.15
Maximum					1.42		1.66
Standard deviation					0.90	0.46	0.18
					1.32		
					2.75		

#The design effect is undefined because the estimate is 100.00.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–28. Student design effects – Asian

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	17.1	1.63	0.98	1465	2.75	1.66
Student born in 1986	DOBIRTHP	68.1	1.72	1.22	1461	1.99	1.41
Student lives with father and mother	BYFCOMP	71.2	1.61	1.18	1465	1.85	1.36
Reading test number-right score	BYTXRIRR	29.9	0.50	0.25	1465	3.84	1.96
Mathematics test number-right score	BYTXMIRR	41.6	0.66	0.32	1465	4.20	2.05
Composite achievement test highest quartile	BYTXCQU	29.9	2.27	1.20	1465	3.61	1.90
Agree or strongly agree that school rules are fair	BYS21B	58.7	2.03	1.33	1380	2.34	1.53
Never was hit the first semester	BYS22E	80.5	1.34	1.06	1398	1.60	1.27
Someone bullied or picked on student at least once	BYS22H	15.6	1.35	0.97	1401	1.94	1.39
High school program is general	SCHPROG	29.6	1.88	1.19	1465	2.49	1.58
Average number of hours spent on homework each week out of school	BYS34B	8.2	0.30	0.19	1427	2.63	1.62
Average number of hours spent on English homework each week out of school	BYS36B	3.3	0.15	0.10	1364	2.25	1.50
Good grades not important	BYS37	0.8	0.26	0.23	1435	1.27	1.13
Never come to class without books	BYS38B	48.9	1.89	1.35	1374	1.97	1.40
Participated in band, orchestra, chorus, or choir	BYS41A	19.7	1.56	1.05	1425	2.18	1.48
Did not participate in a hobby club	BYS41H	84.5	1.41	0.96	1412	2.15	1.47
Use computer as a resource to learn things of interest at least once a week	BYS45C	51.2	1.73	1.36	1357	1.63	1.28
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	51.3	1.99	1.36	1343	2.13	1.46
Being able to give own children better opportunities very important	BYS54G	79.6	1.65	1.09	1376	2.30	1.52
Expect to finish college, but not advanced degree	STEXPECT	33.4	1.89	1.23	1465	2.34	1.53
Plan to continue education right after high school	BYS57	84.2	1.55	0.98	1376	2.48	1.58
Hope to receive an athletic scholarship	BYS61	69.3	2.82	2.06	502	1.88	1.37
At age 30 exp to be a manager	BYOCC30	2.4	0.59	0.43	1274	1.85	1.36
At age 30 exp to be in the military	BYOCC30	0.3	0.16	0.16	1274	1.00	1.00
At age 30 exp to be an operative	BYOCC30	0.3	0.20	0.14	1274	1.94	1.39
At age 30 exp to be a professional (group b)	BYOCC30	24.0	1.64	1.20	1274	1.87	1.37
At age 30 exp to be a technician	BYOCC30	4.3	0.98	0.57	1274	2.97	1.72
At age 30 doesn't know what to be	BYOCC30	40.1	1.74	1.37	1274	1.61	1.27
English is native language	STLANG	36.9	2.01	1.26	1465	2.53	1.59
Among close friends, somewhat or very important that they have a regular job	BYS90K	77.8	2.03	1.30	1020	2.44	1.56
SUMMARY STATISTICS							
Mean							1.49
Minimum							1.00
Median							1.47
Maximum					2.27		2.05
Standard deviation					1.00		0.22
					2.17		
					4.20		
					0.70		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–29. Student design effects – Black

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	59.04	2.30	1.09	2033	4.43	2.10
Student born in 1986	DOBIRTHP	54.52	1.46	1.11	2024	1.73	1.32
Student lives with father and mother	BYFCOMP	31.42	1.28	1.03	2033	1.54	1.24
Reading test number-right score	BYTXRIRR	24.28	0.28	0.18	2033	2.29	1.51
Mathematics test number-right score	BYTXMIRR	29.71	0.35	0.21	2033	2.78	1.67
Composite achievement test highest quartile	BYTXCQU	5.43	0.64	0.50	2033	1.61	1.27
Agree or strongly agree that school rules are fair	BYS21B	40.66	1.68	1.13	1891	2.20	1.48
Never was hit the first semester	BYS22E	76.43	1.18	0.97	1909	1.47	1.21
Someone bullied or picked on student at least once	BYS22H	12.92	0.96	0.77	1921	1.58	1.26
High school program is general	SCHPROG	34.05	1.35	1.05	2033	1.65	1.28
Average number of hours spent on homework each week out of school	BYS34B	4.94	0.16	0.13	1908	1.52	1.23
Average number of hours spent on English homework each week out of school	BYS36B	2.56	0.10	0.08	1725	1.33	1.15
Good grades not important	BYS37	0.50	0.17	0.16	1960	1.12	1.06
Never come to class without books	BYS38B	47.04	1.40	1.15	1871	1.48	1.22
Participated in band, orchestra, chorus, or choir	BYS41A	21.55	1.33	0.94	1925	2.01	1.42
Did not participate in a hobby club	BYS41H	92.18	0.68	0.61	1921	1.21	1.10
Use computer as a resource to learn things of interest at least once a week	BYS45C	44.29	1.41	1.17	1816	1.46	1.21
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	69.43	1.34	1.09	1787	1.51	1.23
Being able to give own children better opportunities very important	BYS54G	88.34	0.88	0.74	1865	1.39	1.18
Expect to finish college, but not advanced degree	STEXPECT	37.34	1.19	1.07	2033	1.23	1.11
Plan to continue education right after high school	BYS57	75.73	1.25	1.01	1786	1.51	1.23
Hope to receive an athletic scholarship	BYS61	81.66	1.40	1.27	928	1.22	1.10
At age 30 exp to be a manager	BYOCC30	2.61	0.53	0.39	1650	1.82	1.35
At age 30 exp to be in the military	BYOCC30	1.02	0.30	0.25	1650	1.44	1.20
At age 30 exp to be an operative	BYOCC30	0.17	0.12	0.10	1650	1.42	1.19
At age 30 exp to be a professional (group b)	BYOCC30	25.16	1.22	1.07	1650	1.31	1.15
At age 30 exp to be a technician	BYOCC30	4.42	0.61	0.51	1650	1.47	1.21
At age 30 doesn't know what to be	BYOCC30	29.36	1.28	1.12	1650	1.30	1.14
English is native language	STLANG	94.39	0.64	0.51	2033	1.56	1.25
Among close friends, somewhat or very important that they have a regular job	BYS90K	85.76	1.28	1.06	1079	1.44	1.20
SUMMARY STATISTICS							
Mean							1.28
Minimum							1.06
Median							1.22
Maximum					1.67		2.10
Standard deviation					1.12	0.63	0.20
					1.49		
					4.43		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–30. Student design effects – Hispanic

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	26.8	2.39	0.94	2234	6.52	2.55
Student born in 1986	DOBIRTHP	55.2	1.26	1.06	2216	1.42	1.19
Student lives with father and mother	BYFCOMP	54.1	1.34	1.05	2234	1.62	1.27
Reading test number-right score	BYTXRIRR	24.6	0.33	0.19	2234	2.81	1.68
Mathematics test number-right score	BYTXMIRR	31.1	0.38	0.23	2234	2.62	1.62
Composite achievement test highest quartile	BYTXCQU	10.6	0.95	0.65	2234	2.12	1.46
Agree or strongly agree that school rules are fair	BYS21B	52.7	1.46	1.09	2099	1.80	1.34
Never was hit the first semester	BYS22E	80.1	1.03	0.87	2101	1.40	1.18
Someone bullied or picked on student at least once	BYS22H	16.7	1.06	0.81	2125	1.71	1.31
High school program is general	SCHPROG	44.1	1.37	1.05	2234	1.70	1.30
Average number of hours spent on homework each week out of school	BYS34B	5.5	0.16	0.13	2118	1.72	1.31
Average number of hours spent on English homework each week out of school	BYS36B	2.9	0.11	0.08	1985	1.76	1.33
Good grades not important	BYS37	1.8	0.33	0.29	2179	1.34	1.16
Never come to class without books	BYS38B	44.7	1.27	1.09	2085	1.36	1.17
Participated in band, orchestra, chorus, or choir	BYS41A	13.0	0.91	0.72	2165	1.58	1.26
Did not participate in a hobby club	BYS41H	92.0	0.64	0.59	2153	1.21	1.10
Use computer as a resource to learn things of interest at least once a week	BYS45C	39.6	1.24	1.08	2051	1.31	1.15
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	59.6	1.36	1.09	2017	1.55	1.25
Being able to give own children better opportunities very important	BYS54G	85.6	0.93	0.77	2071	1.45	1.20
Expect to finish college, but not advanced degree	STEXPECT	35.0	1.30	1.01	2234	1.65	1.28
Plan to continue education right after high school	BYS57	66.5	1.15	1.08	1924	1.14	1.07
Hope to receive an athletic scholarship	BYS61	66.9	1.94	1.63	833	1.42	1.19
At age 30 exp to be a manager	BYOCC30	1.7	0.36	0.30	1855	1.41	1.19
At age 30 exp to be in the military	BYOCC30	0.6	0.22	0.17	1855	1.55	1.24
At age 30 exp to be an operative	BYOCC30	0.2	0.14	0.11	1855	1.49	1.22
At age 30 exp to be a professional (group b)	BYOCC30	18.2	1.02	0.90	1855	1.29	1.14
At age 30 exp to be a technician	BYOCC30	2.1	0.37	0.33	1855	1.25	1.12
At age 30 doesn't know what to be	BYOCC30	42.7	1.48	1.15	1855	1.66	1.29
English is native language	STLANG	47.7	1.93	1.06	2234	3.34	1.83
Among close friends, somewhat or very important that they have a regular job	BYS90K	84.4	1.13	0.98	1374	1.34	1.16
SUMMARY STATISTICS							
Mean							1.32
Minimum							1.07
Median							1.25
Maximum					1.82		2.55
Standard deviation					1.14		0.29
					1.55		
					6.52		
					1.01		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–31. Student design effects – White

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	32.3	1.03	0.50	8757	4.22	2.05
Student born in 1986	DOBIRTHP	58.1	0.69	0.53	8735	1.72	1.31
Student lives with father and mother	BYFCOMP	63.5	0.69	0.51	8757	1.82	1.35
Reading test number-right score	BYTXRIRR	32.1	0.17	0.10	8757	2.99	1.73
Mathematics test number-right score	BYTXMIRR	40.5	0.21	0.12	8757	3.21	1.79
Composite achievement test highest quartile	BYTXCQU	33.9	0.82	0.51	8757	2.62	1.62
Agree or strongly agree that school rules are fair	BYS21B	57.9	0.78	0.54	8289	2.05	1.43
Never was hit the first semester	BYS22E	79.5	0.59	0.44	8340	1.76	1.33
Someone bullied or picked on student at least once	BYS22H	22.7	0.57	0.46	8364	1.57	1.25
High school program is general	SCHPROG	38.6	0.81	0.52	8757	2.45	1.57
Average number of hours spent on homework each week out of school	BYS34B	5.7	0.10	0.06	8605	2.72	1.65
Average number of hours spent on English homework each week out of school	BYS36B	2.3	0.05	0.03	8047	2.51	1.58
Good grades not important	BYS37	1.4	0.15	0.13	8652	1.31	1.14
Never come to class without books	BYS38B	54.8	0.81	0.55	8319	2.23	1.49
Participated in band, orchestra, chorus, or choir	BYS41A	23.9	0.65	0.46	8640	2.02	1.42
Did not participate in a hobby club	BYS41H	90.3	0.47	0.32	8604	2.13	1.46
Use computer as a resource to learn things of interest at least once a week	BYS45C	42.1	0.70	0.54	8265	1.67	1.29
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	46.5	0.71	0.55	8207	1.67	1.29
Being able to give own children better opportunities very important	BYS54G	77.2	0.57	0.46	8485	1.58	1.26
Expect to finish college, but not advanced degree	STEXPECT	35.9	0.60	0.51	8757	1.37	1.17
Plan to continue education right after high school	BYS57	71.7	0.66	0.50	8103	1.74	1.32
Hope to receive an athletic scholarship	BYS61	72.6	1.01	0.75	3499	1.81	1.35
At age 30 exp to be a manager	BYOCC30	1.8	0.18	0.15	7958	1.51	1.23
At age 30 exp to be in the military	BYOCC30	1.1	0.14	0.12	7958	1.41	1.19
At age 30 exp to be an operative	BYOCC30	0.7	0.15	0.10	7958	2.33	1.53
At age 30 exp to be a professional (group b)	BYOCC30	19.8	0.54	0.45	7958	1.49	1.22
At age 30 exp to be a technician	BYOCC30	3.4	0.24	0.20	7958	1.45	1.21
At age 30 doesn't know what to be	BYOCC30	33.6	0.65	0.53	7958	1.52	1.23
English is native language	STLANG	97.0	0.28	0.18	8757	2.33	1.53
Among close friends, somewhat or very important that they have a regular job	BYS90K	76.6	0.70	0.53	6410	1.77	1.33
SUMMARY STATISTICS							
Mean							1.41
Minimum							1.14
Median							1.34
Maximum					2.03		2.05
Standard deviation					1.31		0.21
					1.79		
					4.22		
					0.65		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–32. Student design effects – Multiracial

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	27.2	2.21	1.64	742	1.83	1.35
Student born in 1986	DOBIRTHP	58.2	2.45	1.81	741	1.83	1.35
Student lives with father and mother	BYFCOMP	46.4	2.42	1.83	742	1.74	1.32
Reading test number-right score	BYTXRIRR	29.0	0.48	0.35	742	1.92	1.39
Mathematics test number-right score	BYTXMIRR	36.0	0.55	0.42	742	1.72	1.31
Composite achievement test highest quartile	BYTXCQU	20.5	1.87	1.48	742	1.59	1.26
Agree or strongly agree that school rules are fair	BYS21B	46.8	2.44	1.87	712	1.70	1.30
Never was hit the first semester	BYS22E	69.1	2.13	1.73	718	1.53	1.24
Someone bullied or picked on student at least once	BYS22H	26.3	2.04	1.64	719	1.54	1.24
High school program is general	SCHPROG	40.5	2.23	1.80	742	1.53	1.24
Average number of hours spent on homework each week out of school	BYS34B	5.8	0.27	0.21	720	1.72	1.31
Average number of hours spent on English homework each week out of school	BYS36B	2.4	0.13	0.11	673	1.33	1.15
Good grades not important	BYS37	1.2	0.56	0.40	733	1.94	1.39
Never come to class without books	BYS38B	50.8	2.45	1.87	712	1.71	1.31
Participated in band, orchestra, chorus, or choir	BYS41A	21.3	1.80	1.52	729	1.40	1.18
Did not participate in a hobby club	BYS41H	87.3	1.50	1.24	726	1.47	1.21
Use computer as a resource to learn things of interest at least once a week	BYS45C	42.1	2.22	1.85	710	1.43	1.19
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	60.1	2.55	1.86	695	1.89	1.37
Being able to give own children better opportunities very important	BYS54G	77.8	1.85	1.55	721	1.42	1.19
Expect to finish college, but not advanced degree	STEXPECT	35.0	2.32	1.75	742	1.75	1.32
Plan to continue education right after high school	BYS57	63.4	2.64	1.86	669	2.01	1.42
Hope to receive an athletic scholarship	BYS61	72.0	3.23	2.53	316	1.63	1.28
At age 30 exp to be a manager	BYOCC30	1.9	0.70	0.53	671	1.75	1.32
At age 30 exp to be in the military	BYOCC30	0.8	0.38	0.34	671	1.25	1.12
At age 30 exp to be an operative	BYOCC30	0.6	0.48	0.30	671	2.46	1.57
At age 30 exp to be a professional (group b)	BYOCC30	20.0	1.84	1.55	671	1.41	1.19
At age 30 exp to be a technician	BYOCC30	1.9	0.56	0.53	671	1.12	1.06
At age 30 doesn't know what to be	BYOCC30	30.7	2.20	1.78	671	1.53	1.24
English is native language	STLANG	92.5	1.04	0.97	742	1.15	1.07
Among close friends, somewhat or very important that they have a regular job	BYS90K	83.4	2.08	1.64	518	1.62	1.27
SUMMARY STATISTICS							
Mean							1.27
Minimum							1.06
Median							1.27
Maximum					1.63		1.57
Standard deviation					1.12		0.11
					1.63		
					2.46		
					0.28		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–33. Student design effects – public

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	34.4	0.70	0.43	12039	2.63	1.62
Student born in 1986	DOBIRTHP	57.1	0.58	0.45	11998	1.64	1.28
Student lives with father and mother	BYFCOMP	55.3	0.61	0.45	12039	1.78	1.34
Reading test number-right score	BYTXRIRR	29.1	0.19	0.09	12039	4.57	2.14
Mathematics test number-right score	BYTXMIRR	36.7	0.24	0.11	12039	4.98	2.23
Composite achievement test highest quartile	BYTXCQU	23.6	0.71	0.39	12039	3.41	1.85
Agree or strongly agree that school rules are fair	BYS21B	53.5	0.69	0.47	11363	2.18	1.48
Never was hit the first semester	BYS22E	78.5	0.49	0.38	11435	1.60	1.26
Someone bullied or picked on student at least once	BYS22H	20.4	0.47	0.38	11493	1.56	1.25
High school program is general	SCHPROG	40.1	0.66	0.45	12039	2.18	1.48
Average number of hours spent on homework each week out of school	BYS34B	5.5	0.08	0.05	11634	2.60	1.61
Average number of hours spent on English homework each week out of school	BYS36B	2.4	0.04	0.03	10721	2.01	1.42
Good grades not important	BYS37	1.4	0.12	0.11	11815	1.35	1.16
Never come to class without books	BYS38B	51.2	0.66	0.47	11344	1.98	1.41
Participated in band, orchestra, chorus, or choir	BYS41A	21.2	0.53	0.38	11746	2.01	1.42
Did not participate in a hobby club	BYS41H	91.1	0.35	0.26	11690	1.79	1.34
Use computer as a resource to learn things of interest at least once a week	BYS45C	42.1	0.60	0.47	11203	1.65	1.29
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	53.7	0.60	0.47	11085	1.59	1.26
Being able to give own children better opportunities very important	BYS54G	80.7	0.45	0.37	11447	1.51	1.23
Expect to finish college, but not advanced degree	STEXPECT	35.8	0.49	0.44	12039	1.26	1.12
Plan to continue education right after high school	BYS57	70.4	0.56	0.44	10781	1.61	1.27
Hope to receive an athletic scholarship	BYS61	73.5	0.80	0.65	4565	1.49	1.22
At age 30 exp to be a manager	BYOCC30	1.9	0.15	0.13	10535	1.34	1.16
At age 30 exp to be in the military	BYOCC30	1.0	0.11	0.10	10535	1.34	1.16
At age 30 exp to be an operative	BYOCC30	0.6	0.11	0.07	10535	2.09	1.44
At age 30 exp to be a professional (group b)	BYOCC30	20.0	0.45	0.39	10535	1.36	1.17
At age 30 exp to be a technician	BYOCC30	3.4	0.21	0.18	10535	1.40	1.18
At age 30 doesn't know what to be	BYOCC30	34.6	0.57	0.46	10535	1.51	1.23
English is native language	STLANG	85.5	0.64	0.32	12039	4.02	2.00
Among close friends, somewhat or very important that they have a regular job	BYS90K	79.7	0.59	0.45	8109	1.76	1.33
SUMMARY STATISTICS							
Mean							1.41
Minimum							1.12
Median							1.31
Maximum					2.07		2.23
Standard deviation					1.26		0.29
					1.71		
					4.98		
					0.96		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–34. Student design effects – Catholic

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	24.5	2.13	0.98	1920	4.72	2.17
Student born in 1986	DOBIRTHP	66.1	1.48	1.08	1918	1.86	1.36
Student lives with father and mother	BYFCOMP	75.1	1.23	0.99	1920	1.56	1.25
Reading test number-right score	BYTXRIRR	35.0	0.42	0.19	1920	5.19	2.28
Mathematics test number-right score	BYTXMIRR	43.0	0.48	0.23	1920	4.37	2.09
Composite achievement test highest quartile	BYTXCQU	42.4	2.04	1.13	1920	3.26	1.81
Agree or strongly agree that school rules are fair	BYS21B	57.6	2.06	1.16	1826	3.17	1.78
Never was hit the first semester	BYS22E	80.8	1.48	0.92	1833	2.59	1.61
Someone bullied or picked on student at least once	BYS22H	20.0	0.94	0.93	1841	1.01	1.00
High school program is general	SCHPROG	18.4	1.74	0.88	1920	3.88	1.97
Average number of hours spent on homework each week out of school	BYS34B	8.0	0.23	0.14	1910	2.69	1.64
Average number of hours spent on English homework each week out of school	BYS36B	2.9	0.10	0.06	1877	2.66	1.63
Good grades not important	BYS37	1.2	0.27	0.25	1909	1.20	1.10
Never come to class without books	BYS38B	55.4	1.42	1.16	1834	1.50	1.23
Participated in band, orchestra, chorus, or choir	BYS41A	18.1	1.82	0.88	1906	4.26	2.06
Did not participate in a hobby club	BYS41H	82.9	1.35	0.86	1900	2.44	1.56
Use computer as a resource to learn things of interest at least once a week	BYS45C	48.4	1.97	1.17	1822	2.82	1.68
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	41.7	1.76	1.16	1813	2.31	1.52
Being able to give own children better opportunities very important	BYS54G	77.1	1.18	0.97	1883	1.49	1.22
Expect to finish college, but not advanced degree	STEXPECT	39.0	1.55	1.11	1920	1.95	1.39
Plan to continue education right after high school	BYS57	88.8	1.14	0.73	1878	2.46	1.57
Hope to receive an athletic scholarship	BYS61	75.1	1.62	1.39	969	1.36	1.16
At age 30 exp to be a manager	BYOCC30	2.2	0.43	0.35	1774	1.57	1.25
At age 30 exp to be in the military	BYOCC30	0.7	0.25	0.19	1774	1.68	1.30
At age 30 exp to be an operative	BYOCC30	0.1	0.06	0.06	1774	1.14	1.07
At age 30 exp to be a professional (group b)	BYOCC30	25.3	1.34	1.03	1774	1.70	1.30
At age 30 exp to be a technician	BYOCC30	2.4	0.40	0.36	1774	1.21	1.10
At age 30 doesn't know what to be	BYOCC30	34.2	1.01	1.13	1774	0.80	0.89
English is native language	STLANG	93.9	1.16	0.55	1920	4.50	2.12
Among close friends, somewhat or very important that they have a regular job	BYS90K	72.0	1.54	1.18	1445	1.69	1.30
SUMMARY STATISTICS							
Mean							1.51
Minimum							0.89
Median							1.46
Maximum					2.43		2.28
Standard deviation					0.80		0.38
					2.13		
					5.19		
					1.23		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–35. Student design effects – other private

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	44.2	3.42	1.33	1403	6.67	2.58
Student born in 1986	DOBIRTHP	56.8	1.70	1.33	1392	1.63	1.28
Student lives with father and mother	BYFCOMP	74.0	1.70	1.17	1403	2.10	1.45
Reading test number-right score	BYTXRIRR	34.3	0.68	0.25	1403	7.50	2.74
Mathematics test number-right score	BYTXMIRR	43.1	0.80	0.31	1403	6.79	2.61
Composite achievement test highest quartile	BYTXCQU	43.6	3.26	1.32	1403	6.05	2.46
Agree or strongly agree that school rules are fair	BYS21B	64.9	2.39	1.32	1305	3.27	1.81
Never was hit the first semester	BYS22E	79.9	1.21	1.10	1322	1.21	1.10
Someone bullied or picked on student at least once	BYS22H	15.6	1.29	1.00	1320	1.66	1.29
High school program is general	SCHPROG	22.5	3.39	1.11	1403	9.25	3.04
Average number of hours spent on homework each week out of school	BYS34B	8.4	0.49	0.19	1359	6.74	2.60
Average number of hours spent on English homework each week out of school	BYS36B	3.0	0.14	0.08	1315	2.80	1.67
Good grades not important	BYS37	1.6	0.39	0.34	1362	1.27	1.13
Never come to class without books	BYS38B	56.9	2.11	1.37	1304	2.37	1.54
Participated in band, orchestra, chorus, or choir	BYS41A	33.9	3.61	1.28	1359	7.88	2.81
Did not participate in a hobby club	BYS41H	85.2	2.14	0.97	1352	4.89	2.21
Use computer as a resource to learn things of interest at least once a week	BYS45C	42.6	2.22	1.38	1292	2.60	1.61
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	33.7	2.47	1.33	1265	3.44	1.86
Being able to give own children better opportunities very important	BYS54G	71.5	1.68	1.25	1307	1.81	1.35
Expect to finish college, but not advanced degree	STEXPECT	32.7	1.69	1.25	1403	1.81	1.35
Plan to continue education right after high school	BYS57	79.5	2.22	1.11	1313	3.96	1.99
Hope to receive an athletic scholarship	BYS61	63.5	2.53	1.95	608	1.67	1.29
At age 30 exp to be a manager	BYOCC30	3.1	0.64	0.50	1211	1.65	1.29
At age 30 exp to be in the military	BYOCC30	0.6	0.20	0.22	1211	0.82	0.90
At age 30 exp to be an operative	BYOCC30	0.4	0.25	0.18	1211	1.99	1.41
At age 30 exp to be a professional (group b)	BYOCC30	26.5	1.89	1.27	1211	2.23	1.49
At age 30 exp to be a technician	BYOCC30	2.0	0.46	0.41	1211	1.29	1.14
At age 30 doesn't know what to be	BYOCC30	37.0	1.89	1.39	1211	1.85	1.36
English is native language	STLANG	89.4	1.91	0.82	1403	5.41	2.33
Among close friends, somewhat or very important that they have a regular job	BYS90K	71.1	2.65	1.50	918	3.13	1.77
SUMMARY STATISTICS							
Mean							1.78
Minimum							0.90
Median							1.58
Maximum					3.53		3.04
Standard deviation					0.82		0.60
					2.48		
					9.25		
					2.38		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–36. Student design effects – low socioeconomic status (SES) quartile

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	36.7	1.42	0.80	3635	3.16	1.78
Student born in 1986	DOBIRTHP	49.3	1.05	0.83	3620	1.59	1.26
Student lives with father and mother	BYFCOMP	44.6	1.07	0.82	3635	1.68	1.30
Reading test number-right score	BYTXRIRR	24.4	0.23	0.15	3635	2.50	1.58
Mathematics test number-right score	BYTXMIRR	31.0	0.28	0.18	3635	2.53	1.59
Composite achievement test highest quartile	BYTXCQU	9.3	0.65	0.48	3635	1.85	1.36
Agree or strongly agree that school rules are fair	BYS21B	51.7	1.15	0.86	3397	1.79	1.34
Never was hit the first semester	BYS22E	78.7	0.90	0.70	3413	1.64	1.28
Someone bullied or picked on student at least once	BYS22H	19.6	0.85	0.68	3442	1.59	1.26
High school program is general	SCHPROG	42.8	1.04	0.82	3635	1.59	1.26
Average number of hours spent on homework each week out of school	BYS34B	4.7	0.11	0.09	3459	1.51	1.23
Average number of hours spent on English homework each week out of school	BYS36B	2.4	0.07	0.06	3158	1.42	1.19
Good grades not important	BYS37	1.6	0.23	0.21	3538	1.21	1.10
Never come to class without books	BYS38B	48.6	1.13	0.86	3377	1.73	1.31
Participated in band, orchestra, chorus, or choir	BYS41A	15.6	0.75	0.61	3503	1.51	1.23
Did not participate in a hobby club	BYS41H	93.3	0.50	0.42	3491	1.39	1.18
Use computer as a resource to learn things of interest at least once a week	BYS45C	37.8	1.03	0.84	3306	1.50	1.22
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	62.3	1.02	0.85	3264	1.43	1.20
Being able to give own children better opportunities very important	BYS54G	83.6	0.74	0.64	3384	1.34	1.16
Expect to finish college, but not advanced degree	STEXPECT	33.4	0.89	0.78	3635	1.30	1.14
Plan to continue education right after high school	BYS57	61.4	1.04	0.88	3042	1.39	1.18
Hope to receive an athletic scholarship	BYS61	71.5	1.68	1.30	1203	1.66	1.29
At age 30 exp to be a manager	BYOCC30	1.4	0.27	0.21	3032	1.59	1.26
At age 30 exp to be in the military	BYOCC30	0.6	0.15	0.14	3032	1.11	1.05
At age 30 exp to be an operative	BYOCC30	0.7	0.17	0.15	3032	1.26	1.12
At age 30 exp to be a professional (group b)	BYOCC30	17.6	0.88	0.69	3032	1.61	1.27
At age 30 exp to be a technician	BYOCC30	3.6	0.41	0.34	3032	1.45	1.20
At age 30 doesn't know what to be	BYOCC30	38.4	1.12	0.88	3032	1.61	1.27
English is native language	STLANG	71.2	1.44	0.75	3635	3.68	1.92
Among close friends, somewhat or very important that they have a regular job	BYS90K	83.2	0.96	0.80	2196	1.45	1.21
SUMMARY STATISTICS							
Mean						1.70	1.29
Minimum						1.11	1.05
Median						1.59	1.26
Maximum						3.68	1.92
Standard deviation						0.56	0.19

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–37. Student design effects – middle socioeconomic status (SES) quartiles

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	32.8	0.87	0.55	7388	2.53	1.59
Student born in 1986	DOBIRTHP	58.8	0.73	0.57	7365	1.60	1.27
Student lives with father and mother	BYFCOMP	54.6	0.72	0.58	7388	1.55	1.24
Reading test number-right score	BYTXRIRR	29.3	0.17	0.11	7388	2.38	1.54
Mathematics test number-right score	BYTXMIRR	36.8	0.21	0.13	7388	2.69	1.64
Composite achievement test highest quartile	BYTXCQU	21.2	0.67	0.48	7388	1.97	1.40
Agree or strongly agree that school rules are fair	BYS21B	52.5	0.88	0.60	7020	2.18	1.48
Never was hit the first semester	BYS22E	77.3	0.62	0.50	7067	1.56	1.25
Someone bullied or picked on student at least once	BYS22H	20.3	0.57	0.48	7103	1.41	1.19
High school program is general	SCHPROG	40.8	0.77	0.57	7388	1.83	1.35
Average number of hours spent on homework each week out of school	BYS34B	5.3	0.08	0.06	7180	1.62	1.27
Average number of hours spent on English homework each week out of school	BYS36B	2.4	0.05	0.04	6664	1.85	1.36
Good grades not important	BYS37	1.4	0.18	0.14	7276	1.66	1.29
Never come to class without books	BYS38B	53.1	0.78	0.60	7015	1.70	1.30
Participated in band, orchestra, chorus, or choir	BYS41A	21.6	0.64	0.48	7247	1.73	1.31
Did not participate in a hobby club	BYS41H	91.2	0.39	0.33	7208	1.38	1.18
Use computer as a resource to learn things of interest at least once a week	BYS45C	41.8	0.75	0.59	6944	1.63	1.28
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	53.7	0.76	0.60	6872	1.61	1.27
Being able to give own children better opportunities very important	BYS54G	81.6	0.58	0.46	7047	1.58	1.26
Expect to finish college, but not advanced degree	STEXPECT	37.2	0.70	0.56	7388	1.57	1.25
Plan to continue education right after high school	BYS57	69.4	0.74	0.56	6717	1.73	1.32
Hope to receive an athletic scholarship	BYS61	74.4	1.01	0.79	3013	1.62	1.27
At age 30 exp to be a manager	BYOCC30	1.9	0.21	0.17	6555	1.49	1.22
At age 30 exp to be in the military	BYOCC30	1.1	0.15	0.13	6555	1.35	1.16
At age 30 exp to be an operative	BYOCC30	0.8	0.17	0.11	6555	2.36	1.54
At age 30 exp to be a professional (group b)	BYOCC30	19.8	0.59	0.49	6555	1.42	1.19
At age 30 exp to be a technician	BYOCC30	3.5	0.26	0.23	6555	1.36	1.17
At age 30 doesn't know what to be	BYOCC30	33.4	0.67	0.58	6555	1.30	1.14
English is native language	STLANG	90.3	0.46	0.34	7388	1.83	1.35
Among close friends, somewhat or very important that they have a regular job	BYS90K	81.2	0.69	0.55	5045	1.58	1.26
SUMMARY STATISTICS							
Mean							1.31
Minimum							1.14
Median							1.27
Maximum					1.73		1.64
Standard deviation					1.30		0.13
					1.62		
					2.69		
					0.36		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–38. Student design effects –high socioeconomic status (SES) quartile

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	34.9	1.48	0.72	4339	4.17	2.04
Student born in 1986	DOBIRTHP	62.7	0.96	0.74	4323	1.70	1.30
Student lives with father and mother	BYFCOMP	73.4	0.90	0.67	4339	1.81	1.35
Reading test number-right score	BYTXRIRR	35.0	0.22	0.13	4339	2.70	1.64
Mathematics test number-right score	BYTXMIRR	44.2	0.28	0.17	4339	2.87	1.69
Composite achievement test highest quartile	BYTXCQU	48.4	1.20	0.76	4339	2.49	1.58
Agree or strongly agree that school rules are fair	BYS21B	59.6	1.07	0.77	4077	1.93	1.39
Never was hit the first semester	BYS22E	81.3	0.77	0.61	4110	1.62	1.27
Someone bullied or picked on student at least once	BYS22H	20.6	0.80	0.63	4109	1.61	1.27
High school program is general	SCHPROG	29.9	1.13	0.70	4339	2.62	1.62
Average number of hours spent on homework each week out of school	BYS34B	7.3	0.16	0.10	4264	2.58	1.61
Average number of hours spent on English homework each week out of school	BYS36B	2.7	0.06	0.04	4091	1.94	1.39
Good grades not important	BYS37	1.1	0.21	0.16	4272	1.70	1.31
Never come to class without books	BYS38B	51.6	1.05	0.78	4090	1.80	1.34
Participated in band, orchestra, chorus, or choir	BYS41A	27.1	1.02	0.68	4261	2.24	1.50
Did not participate in a hobby club	BYS41H	86.5	0.79	0.53	4243	2.25	1.50
Use computer as a resource to learn things of interest at least once a week	BYS45C	47.9	0.97	0.78	4067	1.53	1.24
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	40.9	1.12	0.77	4027	2.08	1.44
Being able to give own children better opportunities very important	BYS54G	74.4	0.88	0.67	4206	1.71	1.31
Expect to finish college, but not advanced degree	STEXPECT	35.4	0.91	0.73	4339	1.58	1.26
Plan to continue education right after high school	BYS57	84.2	0.73	0.56	4213	1.69	1.30
Hope to receive an athletic scholarship	BYS61	72.1	1.28	1.02	1926	1.56	1.25
At age 30 exp to be a manager	BYOCC30	2.4	0.31	0.25	3933	1.62	1.27
At age 30 exp to be in the military	BYOCC30	0.9	0.22	0.15	3933	2.27	1.51
At age 30 exp to be an operative	BYOCC30	0.0	0.00	0.01	3933	#	#
At age 30 exp to be a professional (group b)	BYOCC30	24.1	0.88	0.68	3933	1.67	1.29
At age 30 exp to be a technician	BYOCC30	2.8	0.36	0.26	3933	1.84	1.36
At age 30 doesn't know what to be	BYOCC30	33.5	1.12	0.75	3933	2.22	1.49
English is native language	STLANG	92.0	0.55	0.41	4339	1.78	1.34
Among close friends, somewhat or very important that they have a regular job	BYS90K	71.9	1.13	0.79	3231	2.03	1.42
SUMMARY STATISTICS							
Mean							1.39
Minimum							0.36
Median							1.35
Maximum					1.99		2.04
Standard deviation					0.13		0.26
					1.82		
					4.17		
					0.65		

#The design effect is undefined because the estimate is 100.00.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–39. Student design effects –urban

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	35.8	1.27	0.67	5115	3.60	1.90
Student born in 1986	DOBIRTHP	57.5	0.96	0.69	5099	1.91	1.38
Student lives with father and mother	BYFCOMP	52.8	1.24	0.70	5115	3.15	1.78
Reading test number-right score	BYTXRIRR	28.4	0.38	0.14	5115	7.50	2.74
Mathematics test number-right score	BYTXMIRR	35.5	0.48	0.17	5115	8.06	2.84
Composite achievement test highest quartile	BYTXCQU	21.8	1.34	0.58	5115	5.42	2.33
Agree or strongly agree that school rules are fair	BYS21B	52.0	1.18	0.72	4770	2.68	1.64
Never was hit the first semester	BYS22E	80.4	0.78	0.57	4826	1.85	1.36
Someone bullied or picked on student at least once	BYS22H	16.6	0.64	0.54	4841	1.42	1.19
High school program is general	SCHPROG	38.0	1.22	0.68	5115	3.24	1.80
Average number of hours spent on homework each week out of school	BYS34B	6.1	0.14	0.09	4932	2.73	1.65
Average number of hours spent on English homework each week out of school	BYS36B	2.7	0.07	0.05	4692	2.36	1.54
Good grades not important	BYS37	1.2	0.21	0.16	4996	1.74	1.32
Never come to class without books	BYS38B	48.7	0.94	0.72	4767	1.69	1.30
Participated in band, orchestra, chorus, or choir	BYS41A	18.9	0.96	0.56	4951	2.98	1.73
Did not participate in a hobby club	BYS41H	89.3	0.71	0.44	4929	2.64	1.62
Use computer as a resource to learn things of interest at least once a week	BYS45C	44.5	1.22	0.73	4695	2.85	1.69
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	55.7	1.16	0.73	4597	2.49	1.58
Being able to give own children better opportunities very important	BYS54G	82.5	0.78	0.55	4790	2.03	1.43
Expect to finish college, but not advanced degree	STEXPECT	35.8	0.86	0.67	5115	1.67	1.29
Plan to continue education right after high school	BYS57	73.3	0.87	0.65	4682	1.83	1.35
Hope to receive an athletic scholarship	BYS61	73.3	1.34	0.98	2035	1.87	1.37
At age 30 exp to be a manager	BYOCC30	2.6	0.31	0.24	4379	1.70	1.30
At age 30 exp to be in the military	BYOCC30	0.9	0.19	0.14	4379	1.86	1.36
At age 30 exp to be an operative	BYOCC30	0.1	0.08	0.06	4379	2.05	1.43
At age 30 exp to be a professional (group b)	BYOCC30	22.8	0.81	0.63	4379	1.63	1.28
At age 30 exp to be a technician	BYOCC30	3.3	0.38	0.27	4379	1.97	1.41
At age 30 doesn't know what to be	BYOCC30	35.3	0.96	0.72	4379	1.77	1.33
English is native language	STLANG	78.7	1.50	0.57	5115	6.90	2.63
Among close friends, somewhat or very important that they have a regular job	BYS90K	78.1	1.21	0.72	3322	2.83	1.68
SUMMARY STATISTICS							
Mean							1.64
Minimum							1.19
Median							1.48
Maximum					2.88		2.84
Standard deviation					1.42		0.44
					2.20		
					8.06		
					1.76		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–40. Student design effects – suburban

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	29.7	0.89	0.53	7399	2.78	1.67
Student born in 1986	DOBIRTHP	58.3	0.80	0.57	7369	1.95	1.40
Student lives with father and mother	BYFCOMP	59.4	0.74	0.57	7399	1.66	1.29
Reading test number-right score	BYTXRIRR	29.9	0.24	0.11	7399	4.39	2.10
Mathematics test number-right score	BYTXMIRR	38.0	0.30	0.14	7399	4.75	2.18
Composite achievement test highest quartile	BYTXCQU	26.9	0.96	0.52	7399	3.44	1.86
Agree or strongly agree that school rules are fair	BYS21B	54.9	0.93	0.59	7030	2.45	1.57
Never was hit the first semester	BYS22E	78.6	0.58	0.49	7052	1.40	1.19
Someone bullied or picked on student at least once	BYS22H	21.1	0.64	0.48	7093	1.73	1.32
High school program is general	SCHPROG	38.2	0.87	0.57	7399	2.37	1.54
Average number of hours spent on homework each week out of school	BYS34B	5.7	0.12	0.07	7202	3.13	1.77
Average number of hours spent on English homework each week out of school	BYS36B	2.4	0.05	0.04	6687	2.23	1.49
Good grades not important	BYS37	1.4	0.17	0.14	7279	1.58	1.26
Never come to class without books	BYS38B	52.0	0.91	0.60	7014	2.34	1.53
Participated in band, orchestra, chorus, or choir	BYS41A	21.9	0.70	0.49	7262	2.05	1.43
Did not participate in a hobby club	BYS41H	90.9	0.46	0.34	7223	1.84	1.36
Use computer as a resource to learn things of interest at least once a week	BYS45C	42.0	0.73	0.59	6952	1.53	1.24
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	50.6	0.76	0.60	6911	1.58	1.26
Being able to give own children better opportunities very important	BYS54G	79.4	0.61	0.48	7112	1.63	1.28
Expect to finish college, but not advanced degree	STEXPECT	35.8	0.63	0.56	7399	1.28	1.13
Plan to continue education right after high school	BYS57	72.1	0.73	0.55	6748	1.78	1.33
Hope to receive an athletic scholarship	BYS61	72.8	1.13	0.82	2973	1.91	1.38
At age 30 exp to be a manager	BYOCC30	1.7	0.19	0.16	6593	1.44	1.20
At age 30 exp to be in the military	BYOCC30	0.9	0.15	0.12	6593	1.54	1.24
At age 30 exp to be an operative	BYOCC30	0.6	0.13	0.09	6593	1.97	1.40
At age 30 exp to be a professional (group b)	BYOCC30	19.6	0.61	0.49	6593	1.54	1.24
At age 30 exp to be a technician	BYOCC30	3.2	0.26	0.22	6593	1.47	1.21
At age 30 doesn't know what to be	BYOCC30	34.5	0.81	0.59	6593	1.89	1.38
English is native language	STLANG	87.0	0.70	0.39	7399	3.23	1.80
Among close friends, somewhat or very important that they have a regular job	BYS90K	79.4	0.73	0.56	5244	1.71	1.31
SUMMARY STATISTICS							
Mean							1.44
Minimum							1.13
Median							1.37
Maximum					2.15		2.18
Standard deviation					1.28		0.27
					1.87		
					4.75		
					0.86		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.

Table K–41. Student design effects – rural

Survey item (or composite variable)	Variable	Estimate	Design standard error	Simple random sample standard error	N	DEFF	DEFT
Student lives in the South	BYREGION	43.9	1.66	0.93	2848	3.17	1.78
Student born in 1986	DOBIRTHP	55.1	1.11	0.93	2840	1.41	1.19
Student lives with father and mother	BYFCOMP	56.5	1.14	0.93	2848	1.51	1.23
Reading test number-right score	BYTXRIRR	30.1	0.33	0.18	2848	3.35	1.83
Mathematics test number-right score	BYTXMIRR	37.9	0.42	0.21	2848	3.80	1.95
Composite achievement test highest quartile	BYTXCQU	25.4	1.29	0.82	2848	2.49	1.58
Agree or strongly agree that school rules are fair	BYS21B	55.2	1.42	0.96	2694	2.21	1.49
Never was hit the first semester	BYS22E	76.2	1.33	0.82	2712	2.66	1.63
Someone bullied or picked on student at least once	BYS22H	23.2	1.13	0.81	2720	1.94	1.39
High school program is general	SCHPROG	40.2	1.34	0.92	2848	2.12	1.46
Average number of hours spent on homework each week out of school	BYS34B	5.1	0.17	0.10	2769	2.91	1.71
Average number of hours spent on English homework each week out of school	BYS36B	2.2	0.09	0.06	2534	2.37	1.54
Good grades not important	BYS37	1.5	0.23	0.23	2811	1.00	1.00
Never come to class without books	BYS38B	55.1	1.51	0.96	2701	2.50	1.58
Participated in band, orchestra, chorus, or choir	BYS41A	24.3	1.27	0.81	2798	2.45	1.56
Did not participate in a hobby club	BYS41H	91.4	0.67	0.53	2790	1.60	1.27
Use computer as a resource to learn things of interest at least once a week	BYS45C	40.0	1.14	0.95	2670	1.44	1.20
Watch more than 2 hrs of TV or videotapes/DVDs per weekday	BYS48A	52.8	1.28	0.97	2655	1.74	1.32
Being able to give own children better opportunities very important	BYS54G	79.2	0.90	0.78	2735	1.35	1.16
Expect to finish college, but not advanced degree	STEXPECT	35.9	1.07	0.90	2848	1.41	1.19
Plan to continue education right after high school	BYS57	67.4	1.35	0.93	2542	2.10	1.45
Hope to receive an athletic scholarship	BYS61	74.0	1.22	1.30	1134	0.88	0.94
At age 30 exp to be a manager	BYOCC30	1.6	0.27	0.25	2548	1.22	1.11
At age 30 exp to be in the military	BYOCC30	1.1	0.22	0.20	2548	1.12	1.06
At age 30 exp to be an operative	BYOCC30	1.1	0.36	0.21	2548	2.89	1.70
At age 30 exp to be a professional (group b)	BYOCC30	19.0	0.91	0.78	2548	1.38	1.18
At age 30 exp to be a technician	BYOCC30	3.8	0.44	0.38	2548	1.36	1.17
At age 30 doesn't know what to be	BYOCC30	33.9	0.93	0.94	2548	0.98	0.99
English is native language	STLANG	94.5	0.47	0.43	2848	1.22	1.10
Among close friends, somewhat or very important that they have a regular job	BYS90K	79.5	1.15	0.93	1906	1.54	1.24
SUMMARY STATISTICS							
Mean							1.37
Minimum							0.94
Median							1.29
Maximum					1.94		1.95
Standard deviation					0.88		0.27
					1.67		
					3.80		
					0.78		

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), Public-use data file.