Education Longitudinal Study of 2002 (ELS:2002/12) Third Follow-Up Field Test Report

Working Paper Series

The Working Paper Series was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

NCES 2012-03

U.S. DEPARTMENT OF EDUCATION

This page left intentionally blank.

Education Longitudinal Study of 2002 (ELS:2002/12) Third Follow-Up Field Test Report

Working Paper Series

May 2012

Steven J. Ingels
Daniel J. Pratt
Donna M. Jewell
Tiffany Mattox
Ben Dalton
Jeffrey Rosen
Erich Lauff
Jason Hill
RTI International

Elise Christopher
Project Officer
National Center for Education Statistics

NCES 2012-03 U.S. DEPARTMENT OF EDUCATION

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Statistics
Jack Buckley
Commissioner

Elementary/Secondary & Libraries Studies Division
Jeffrey A. Owings
Associate 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, and 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, IES, U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

May 2012

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

Suggested Citation

Ingels, S.J., Pratt, D.J., Jewell, D.M., Mattox, T., Dalton, B., Rosen, J., Lauff, E., and Hill, J. (2012). Education Longitudinal Study of 2002 (ELS:2002/12) Third Follow-Up Field Test Report (NCES 2012-03). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC.

Content Contact
Elise Christopher
(202) 502–7899
Elise.christopher@ed.gov

Foreword

In addition to official NCES publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series. Copies of working papers can be downloaded as PDF files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch).

Marilyn M. Seastrom
Chief Mathematical Statistician
Statistical Standards Program

Chen-Su Chen
Mathematical Statistician
Statistical Standards Program

This page intentionally left blank.

Acknowledgments

We thank all the individuals who participated in any or all four rounds of the Education Longitudinal Study of 2002 (ELS:2002), including the thousands of students, parents, and school personnel who generously gave their time and approval to provide the data that the study reports.

The authors of this report also thank the many individuals who assisted in the planning of the ELS:2002 third follow-up. We are particularly indebted to the ELS:2002 third follow-up Technical Review Panel, which met to review plans for the study, helped refine them, and provided important suggestions to help guide development of the instrumentation. The following individuals serve as the nonfederal members of the ELS:2002 third follow-up Technical Review Panel: Sara Goldrick-Rab, Robert Gonyea, Donald Heller, Robert Lent, Amaury Nora, Randall Olsen, Aaron Pallas, Kent Phillippe, Michael Shanahan, and Marvin Titus.

This page intentionally left blank.

Table of Contents

Foreword iii

Acknowledgments v

Table of Contents vii

List of Tables ix

List of Figures xi

Chapter 1. Introduction 1

1.1 Historical Background: NCES Secondary Longitudinal Studies Program 1
1.1.1 National Longitudinal Study of the High School Class of 1972 2
1.1.2 High School and Beyond 4
1.1.3 National Education Longitudinal Study of 1988 5
1.1.4 High School Longitudinal Study of 2009 5
1.2 Education Longitudinal Study of 2002 6
1.2.1 ELS:2002 Research and Policy Issues 6
1.2.2 Overview of Data Collections 7
1.2.3 ELS:2002 Study Design 8
Base Year (2002) 9
First Follow-up (2004) 10
Second Follow-up (2006) 11
Third Follow-up Field Test (2011) and Main Study (2012) 12

Chapter 2. Field Test Survey Design and Preparation 15

2.1 ELS:2002 Third Follow-up Data Collection Instrumentation 15
2.1.1 Instrument Development Goals and Constraints 15
2.1.2 Instrument Development Process 16
2.2 Pre-Data Collection Activities 17
2.2.1 Training of Data Collectors 17
2.2.2 Locating and Tracing Activities 18

Chapter 3. Data Collection Procedures and Results 21

3.1 Web/CATI Data Collection Procedures 21
3.1.1 Study Website and Help Desk 21
3.1.2 Interviewing 22
3.2 Interview Data Collection Outcomes 25
3.3 Results of the Propensity Modeling Experiment 33

Chapter 4. Field Test Questionnaire Timing and Data Quality 43

4.1 Questionnaire Timing and Item Analyses 43
4.1.1 Questionnaire Timing 43

4.1.2 Item Timing 44
4.1.3 Item Response Distribution 46
4.2 Items With "Please Specify" Text 50
4.2.1 Methodological Probes 50
4.2.2 "Please Specify" Responses at End of Series 52
4.3 Reliability Reinterview: Design and Results 52
4.4 Reliability Analysis for Scales 56

Chapter 5. Survey Control Systems and Data Processing 59

5.1 System Design, Development, and Testing 59
5.1.1 Integrated Management System 60
5.1.2 Survey Control System 60
5.1.3 Hatteras Survey Engine and Survey Editor 60
5.1.4 Web Survey 61
5.1.5 Computer-Assisted Telephone Interview Case Management System 61
5.1.6 ELS:2002 Sample-Member Website 62
5.2 Data Processing and File Preparation 62

Chapter 6. Summary of Recommendations for the Full-Scale Study 63

6.1 Recommendations for Pre-Data Collection Procedures 63
6.2 Recommendations for Data Collection 64
6.3 Recommendations for the Study Instrument 64
6.4 Recommendations for the Survey Control Systems and Data Processing 65

References 67

Appendixes

A: ELS:2002 Technical Review Panel Summary A-1

B: ELS:2002 Data Collection Materials B-1

C: ELS:2002 Third Follow-up Field Test Questionnaire C-1

D: ELS:2002 Cognitive Testing Reports, 2010 and 2011 ....D-1

E: ELS:2002 Scale Reliability Analyses E-2

F: ELS:2002 Third Follow-up Field Test Codebooks F-1

List of Tables

Number **Page**

1. Student locating snapshot, by dropout status and second follow-up response status: 2011 27
2. Batch processing confirmation/update rates, by tracing source for 2010 and 2011: 2011 28
3. Address update participation by incentive group: 2011 29
4. Interview completion rate, by address update and incentive: 2011 29
5. Sample members requiring intensive tracing procedures, by dropout status and second follow-up response status: 2011 30
6. Interview completion rate, by address update and incentive: 2011 30
7. Distribution of completed interviews, by data collection phase: 2011 31
8. Number of cases and percentage of completed interviews by data collection phase: 2011 31
9. Number and average of calls, by dropout status and prior- and current-round response status and mode of interview: 2011 32
10. Refusal and refusal conversion rates, by dropout status and prior-round response status: 2011 33
11. Distribution of third follow-up propensity cases by prior response status: 2011 36
12. Number of cases and percentage of completed interviews by data collection phase: 2011 37
13. ELS:2002 field test treatment schedule 37
14. Response rates by propensity and experimental group 38
15. Estimates of select survey variables by propensity group ($p < .05$): 2011 39
16. Unit-level bias analyses: 2011 40
17. Questionnaire timing of ELS:2002 third follow-up field test questionnaire, by response mode and questionnaire section: 2011 44
18. Low-frequency responses to ELS:2002 third follow-up field test questionnaire items: 2011 47
19. Items with “please specify” text: Methodological probes and items with an “other” option: ELS:2002 third follow-up field test: 2011 50
20. Interview-reinterview agreement for categorical items on the ELS:2002 third follow-up field test reliability reinterview: 2011 53

21. Interview-reinterview agreement for continuous items on the ELS:2002 third follow-up field test reliability reinterview: 2011 55

22. Reliability analysis of questionnaire scales: Standardized alpha..... 57

List of Figures

Number **Page**

1. Longitudinal design for the NCES high school cohorts: 1972–2016 3
2. ELS:2002/12 tracing procedures: 2011 19
3. ELS:2002/12 website home page: 2011 22
4. Overall locating and interviewing rates at the time work stopped..... 26

This page left intentionally blank.

Chapter 1. Introduction

The Education Longitudinal Study of 2002 (ELS:2002) is conducted by RTI International—a not-for-profit university-affiliated research organization—for the National Center for Education Statistics (NCES), a part of the Institute of Education Sciences in the U.S. Department of Education. This report describes the methodologies and results of the third follow-up (ELS:2002/12) field test which was conducted in the summer of 2011.The field test report is divided into six chapters:

- Chapter 1: Introduction
- Chapter 2: Field Test Survey Design and Preparation
- Chapter 3: Data Collection Procedures and Results
- Chapter 4: Field Test Questionnaire Timing and Data Quality
- Chapter 5: Survey Control Systems and Data Processing
- Chapter 6: Summary of Recommendations for the Full-scale Study

There also are six appendixes, which include proceedings of the two Technical Review Panel meetings (appendix A), recruitment materials (field test letters and scripts) (appendix B), a hardcopy facsimile of the electronic questionnaire (appendix C), a report on two sets of cognitive interviews (appendix D), supplemental data on psychological scales (appendix E), and a codebook with response frequencies (appendix F). Chapter 1 addresses two broad areas: it provides information on the historical background of ELS:2002—the predecessor and successor NCES Secondary Longitudinal Studies—and an overview of the design and purposes of ELS:2002.

1.1 Historical Background: NCES Secondary 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 secondary school students, NCES instituted the Secondary 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 educational careers, and the personal, familial, social, institutional, and cultural factors that may affect that development.

NCES is 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 with the Institute of Education Sciences, in which NCES is now housed.

The Secondary Longitudinal Studies program consists of three completed and two ongoing studies. Completed studies are 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). The two continuing studies are ELS:2002 and the High School Longitudinal Study of 2009 (HSLS:09).

For ELS:2002—the fourth longitudinal study in the series—base-year, first, and second follow-up data are now available; the ELS:2002 third follow-up full-scale study will take place in the second half of 2012. For HSLS:09—the fifth study in the series—base-year data have now been released, and first follow-up data will be collected in the spring of 2012. Taken together, these studies describe the educational experiences of students from five decades—the 1970s, 1980s, 1990s, 2000s, and 2010s—and also provide bases for further understanding the correlates of educational success in the United States.

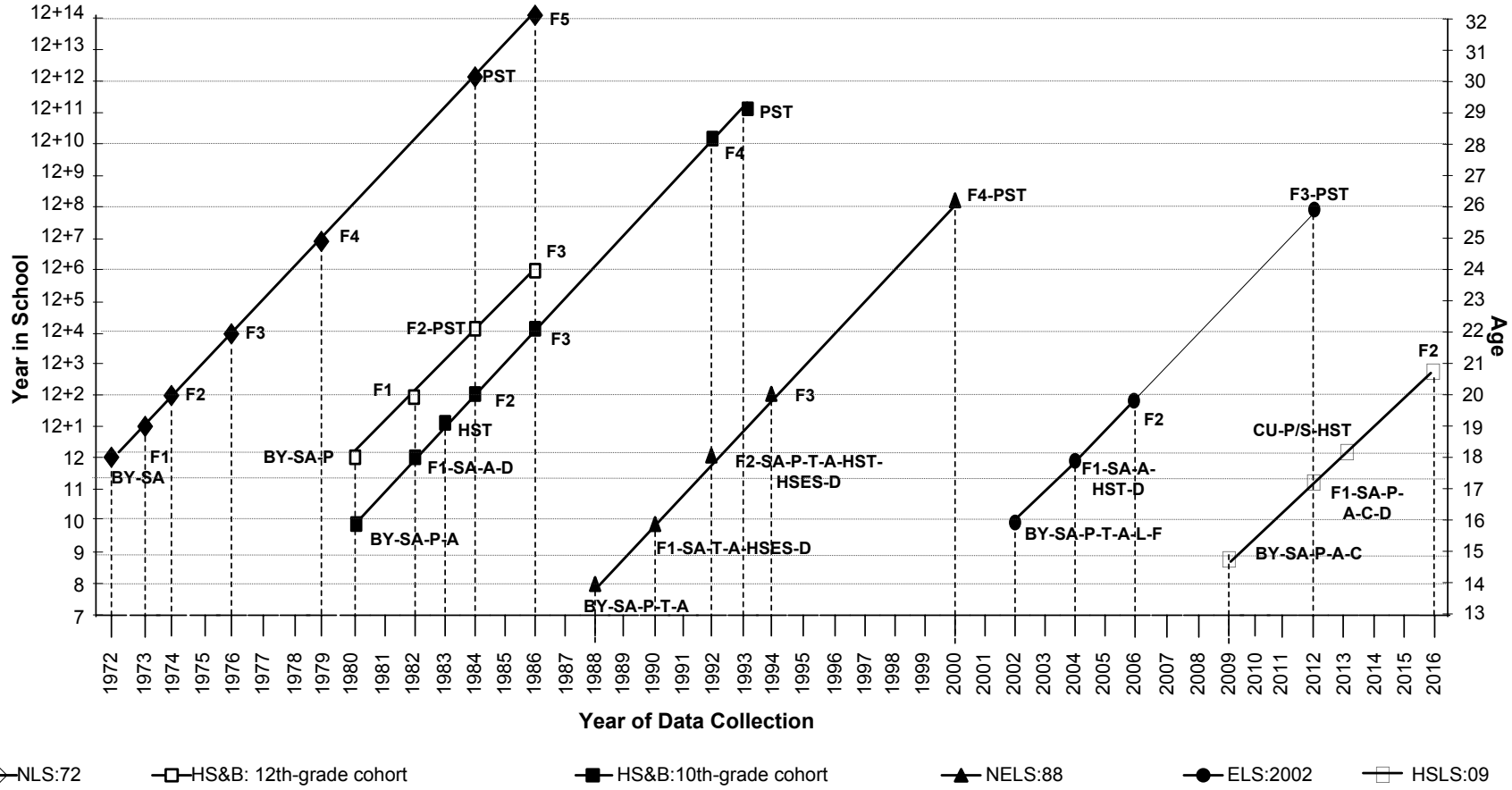
Figure 1 includes a temporal representation of these five longitudinal education studies and highlights their component and comparison points for the time frame 1972–2016.

1.1.1 National Longitudinal Study of the High School Class of 1972

The Secondary Longitudinal Studies program began more than 40 years ago with the implementation of startup activities for NLS:72.[1] NLS:72 was designed to provide longitudinal data for educational policymakers and researchers to 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. School counselors also completed a questionnaire. No parent survey was conducted. However, postsecondary education transcripts were collected in 1984 from the institutions attended by sample members. 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.

[1] For documentation of the NLS:72 project, see Riccobono et al. (1981) and Tourangeau et al. (1987).

Figure 1. Longitudinal design for the NCES high school cohorts: 1972–2016



NLS:72=National Longitudinal Study of the High School Class of 1972
HS&B=High School and Beyond: 1980
NELS:88=National Education Longitudinal Study of 1988
ELS:2002=Education Longitudinal Study of 2002
HSLS:09=High School Longitudinal Study of 2009

BY=Base-year data collection
F1=1st follow-up data collection
F2=2nd follow-up data collection
F3=3rd follow-up data collection
F4=4th follow-up data collection
F5=5th follow-up data collection

P=Parent survey
T=Teacher survey
A=Administrator survey
L=Library/media center survey
F=Facilities checklist
CU=College update
HST=High school transcript
SA=Student assessment

PST=Postsecondary transcript
C=Counselor questionnaire
P/S=Parent or student
HSES=HS effectiveness study
D=Dropout survey

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.

The ELS:2002 sophomore cohort has no parallel in NLS:72, but the ELS:2002 and NLS:72 senior cohorts can be compared.

1.1.2 High School and Beyond

The second in the series of NCES secondary 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 those 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.[2] In addition, to better understand the school and home contexts of 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 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 (1980 to 1986 for the senior cohort, 1982 to 1993 for the sophomore cohort).

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.

[2] For a summation of the HS&B sophomore cohort study, see Zahs et al. (1995). For further information on HS&B, see the NCES website: http://nces.ed.gov/surveys/hsb/.

ELS:2002 data can be compared with the HS&B senior (1980) and sophomore (1980, 1982) cohorts; and indeed, the overall ELS:2002 design is modeled on the HS&B sophomore cohort (both begin at 10th grade and returned to [mostly] the same schools 2 years later).

1.1.3 National Education Longitudinal Study of 1988

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 baseline measure of their achievement and status, prior to their entry into high school. NELS:88 is 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.

Data collection for NELS:88 was initiated with the 8th-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 cohort was also surveyed twice after their scheduled high school graduation, in 1994 and 2000.[3] High school transcripts were collected in the autumn of 1992 and postsecondary transcripts in the autumn of 2000. Through a process of sample freshening, NELS:88 offers three nationally representative cohorts of students: spring-term 8th-, 10th-, and 12th-graders. The NELS:88 10th- and 12th-grade cohorts can be compared with the ELS:2002 10th- and 12th-grade cohorts at grades 10 and 12, and the modal 2 years out of high school, and 8 years out of high school data collections as well as high school (and postsecondary) transcripts.

1.1.4 High School Longitudinal Study of 2009

HSLS:09 is the successor study to ELS:2002. The HSLS:09 base year took place in the fall term of the 2009–10 school year, with a randomly selected sample of fall-term 9th-graders in more than 900 public and private high schools with both a 9th and an 11th grade.[4] Students took a mathematics assessment and survey online. (Also, a small number completed the survey in a telephone interview.) Students' parents, principals, and mathematics and science teachers as well as the school's lead counselor completed surveys on the telephone or on the Web.

The first follow-up of HSLS:09 will take place in the spring of 2012 when most sample members will be in 11th grade. A postsecondary status update (called the "college update," but broader in that it encompasses labor market participation) will take place in the summer/early fall

[3] The entire compass of NELS:88, from its baseline through its final follow-up in 2000, is described in Curtin et al. (2002). NCES keeps an updated version of the NELS:88 bibliography on its website. The bibliography encompasses both project documentation and research articles, monographs, dissertations, and paper presentations employing NELS:88 as well as ELS:2002 data (see http://nces.ed.gov/bibliography).

[4] Types of schools that were excluded from the sample based on the HSLS:09 eligibility definition are described as part of the discussion of the target population in the *HSLS:09 Base-Year Data File Documentation* (see chapter 3, section 3.2.1), Ingels et al. (2011).

of 2013, which will be a survey of the cohort's postsecondary plans and decisions. High school transcripts will be collected in the 2013–14 academic year, and a second follow-up will take place in 2015, when most sample members will be 2 years beyond high school graduation, or in 2016, when most sample members will be 3 years beyond high school graduation. Further follow-ups are contemplated, to at least 2021.

The core research questions for HSLS:09 explore secondary to postsecondary transition plans and the evolution of those plans; the paths into and out of science, technology, engineering, and mathematics; and the educational and social experiences that affect these shifts.

Specific grade-based comparisons between HSLS:09 and ELS:2002 will not be possible, because HSLS:09 is a study of entering 9th-graders (a cohort that will not be freshened at later grades) and ELS:2002 is a study of spring-term (2002) high school sophomores and spring-term (2004) high school seniors.

1.2 Education Longitudinal Study of 2002

1.2.1 ELS:2002 Research and Policy Issues

Apart from helping to describe the status of high school students and their schools, ELS:2002 provides information to help address a number of key policy and research questions. The study is intended to produce a comprehensive dataset for the development and evaluation of education policy at all government levels. Part of its aim is to inform decision makers, education practitioners, and parents about the changes in the operation of the educational system over time. Issues that can be addressed with data collected in the high school years include the following:

- the process of dropping out of high school;
- students' academic growth in mathematics;
- the relationship between family background and the home education support system, and students' high school outcomes;
- the relationship between coursetaking choices and success in the high school years (and thereafter);
- the distribution of educational opportunities as registered in the distinctive school experiences and performance of students from various subgroups, including the following:
    - 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.

Now that most ELS:2002 students have completed high school, a new set of issues is being examined with the help of data collected in 2006. These issues include the following:

- the later educational and labor market activities of high school dropouts; and
- access to and choice of postsecondary educational institutions.

The third follow-up (2012) questionnaire and postsecondary academic transcripts data will support further investigations, such as the following:

- persistence in attaining postsecondary educational goals;
- rate of progress through the postsecondary curriculum;
- degree attainment;
- barriers to persistence and attainment;
- entry of new postsecondary graduates into the workforce;
- social and economic rate of return on education to both the individual and society; and
- adult roles, such as family formation and civic participation.

These various 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 (2002), and seniors (2004);
- longitudinal analyses (including examination of life course changes);
- cross-cohort comparisons (for example, comparisons of high school seniors in 1972, 1980, 1992, and 2004; comparisons of high school seniors 8 years beyond high school in 2000 and 2012); and
- international comparisons: U.S. 15-year-olds to 15-year-olds in other nations, including longitudinal outcomes for the United States that can be related to scale scores in mathematics and reading from the Program for International Student Assessment (PISA).

1.2.2 Overview of Data Collections

ELS:2002 represents a major longitudinal effort designed to provide trend data about critical transitions experienced by students as they proceed through high school and into postsecondary education or their careers. The 2002 sophomore cohort and the distinct but overlapping 2004 senior cohort are being followed to collect policy-relevant data about educational processes and outcomes in the transition to adulthood. These data pertain especially

to student learning, predictors of dropping out, and high school correlates of students' access to and persistence and attainment in postsecondary education, and their entry into the workforce.

In the spring term of 2002, the base year of the study, high school sophomores were assessed in reading and mathematics and surveyed in a national sample of high schools with 10th grades. Their parents, teachers, principals, and librarians were surveyed as well.

In the first follow-up 2 years later (spring term 2004), base-year students who remained in their base-year schools were resurveyed and tested in mathematics, along with a freshening sample that makes the study representative of spring-term 2004 high school seniors nationwide. Students who had transferred to a different school, had switched to a homeschool environment, graduated early, or who had dropped out were administered a questionnaire. Also in the first follow-up, academic transcripts were requested for all students who participated in either the base year or the first follow-up. The transcripts normally[5] cover 4 years of coursework—for students who were seniors in 2004, typically 9th through 12th grade. School course offerings information for the base-year schools was also collected.

The second follow-up, in 2006, took place 2 years after modal graduation from high school, and consisted of an electronically administered questionnaire, supplemented by matches to various educational and postsecondary financial aid records. The administration of the third and final follow-up will take place in the summer and fall of 2012, 6 years after modal graduation from high school, followed by a postsecondary educational transcripts collection.

1.2.3 ELS:2002 Study Design

The transition through high school and beyond into postsecondary institutions and the labor market is both complex (youth may follow many different paths) and prolonged (it takes place over a period of years). The complexity and time frame for this transition make longitudinal approaches especially appropriate.

In addition to achievement tests and questionnaires, information from or linkages to external data sources has been integrated into the ELS:2002 dataset. These external sources include the decennial Census (2000), NCES school databases such as the Common Core of Data (CCD) and Private School Survey (PSS), as well as post-high school institutional information such as the NCES Integrated Postsecondary Education Data System (IPEDS). Additional sources that have been drawn on or linked to include student application and loan information, including the Free Application for Federal Student Aid (FAFSA), and various sources of test scores (SAT, ACT, and the GED testing program) and the National Student Loan Data System (NSLDS).

With the addition of postsecondary data in the 2006 second follow-up (modal age = 20) and the 2012 third follow-up (modal age = 26), ELS:2002 greatly enlarges its ability to connect

[5]Although the goal is to collect four complete years of transcript data, this goal is not achieved for students who fall behind modal grade progression—the transcripts collected for students who were held back, or who dropped out, are incomplete.

high school antecedents to later outcomes. For students who continue 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 workforce (whether as dropouts or high school graduates), ELS:2002 can help 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)

The ELS:2002 base year achieved the following:

- completed the baseline survey of high school sophomores in spring term 2002;
- administered achievement tests in reading and mathematics;
- completed surveys 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 752 participating schools and 15,362 participating students—schools are the first-stage unit of selection, with sophomores randomly selected within schools;
- oversampled Asian[6] students and private schools;
- designed linkages with PISA (reading in 2000 and math in 2003) and National assessment of Educational Progress (NAEP) (math in 2005); and
- scored reporting linkages to the prior longitudinal studies.

The ELS:2002 base-year study 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 was reassessed 2 years hence, so that achievement gain over the last 2 years of high school could be measured and related to school

[6] Except where indicated otherwise, race/ethnicity is reported as follows: Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native. All race categories exclude individuals of Hispanic or Latino origin.

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 attitudes 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 sampled student, an English teacher and a mathematics teacher were 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, staffing, technological resources, collections and expenditures, and scheduling and transactions. Finally, the facilities checklist was a brief observational form completed for each school, which collected information about the condition of school buildings and facilities.

First Follow-up (2004)

The ELS:2002 first follow-up involved the following:

- the survey returned to the same schools but separately followed transfer students and surveyed them outside of school;
- the survey freshened for a spring-term 2004 senior cohort;
- most sample members were seniors, but some were dropouts or in other grades (early graduates or retained in an earlier grade);
- survey instruments included a student questionnaire (different versions for students who remained in the base-year school, transferred to a new school, completed high school early, or were homeschooled), a dropout questionnaire, an assessment in mathematics, and a school administrator questionnaire; and
- there was a high school transcript component in 2004–05 (coursetaking records at the student level for grades 9–12) and a course offerings component at the school level.

The basis for the sampling frame for the first follow-up was the sample of schools and students studied in the ELS:2002 base year. There were two overlapping but conceptually different target student populations, or populations of inferential interest, for the first follow-up. One population (the ELS:2002 sophomore cohort) consists of those students who were enrolled in the 10th grade in the spring term of 2002. The other population (the ELS:2002 senior cohort) comprises those students who were enrolled in the 12th grade in the spring term of 2004. The former population includes students who dropped out of school between 10th and 12th grades, students who graduated early, students who went from a school setting to a homeschooling setting, and students who fell behind the modal grade progression of their peers (e.g., students who repeated a grade and were 11th-graders in spring 2004). Because of these two target

populations and the major analytical subgroups, the full-scale sample encompasses the following types of students in the spring of 2004:

- ELS:2002 base-year sophomores enrolled (in either the 12th grade or some other grade) in the school in which they were originally sampled;
- ELS:2002 base-year sophomores who dropped out of school prior to first follow-up (2004) data collection;
- ELS:2002 base-year sophomores who finished high school early, including those who graduated from high school early and those who did not graduate because they achieved alternative certification (e.g., exam-certified equivalency such as a GED);
- ELS:2002 base-year sophomores who transferred out of the school in which they were originally sampled (including homeschooled students);
- ELS:2002 base-year sample sophomores who were deemed unable to participate directly during the base year as a result of severe disability or insufficient command of the English language such that they could not complete a questionnaire; and
- students at the ELS:2002 base-year sample school who were enrolled in the 12th grade in the spring term of 2004 but who were not in 10th grade in the United States during the 2001–02 school year. In spring term 2002, such students may have been out of the country, been enrolled in school in the United States in a grade other than 10th, had an extended illness or injury, been homeschooled, been institutionalized, or temporarily dropped out of school. These students comprised the first follow-up "freshening" sample.

Although all groups in the sample as categorized above were eligible to complete a questionnaire, different instruments were tailored to different study populations. The guiding intuition was to provide a core of items to which all sample members would respond, supplemented by items specific to the circumstances of a particular group (such as dropouts, for example, for whom questions about their current school situation would not be relevant).

Second Follow-up (2006)

The ELS:2002 second follow-up had the following characteristics:

- A post–high school follow-up employed a web-based instrument for self-administration, a computer-assisted telephone interview (CATI), and a computer-assisted personal interview (CAPI).
- Two years after the ELS:2002 cohorts' modal high school graduation, the second follow-up captures five distinct (but potentially overlapping) groups:
    - late completers and noncompleters of high school;
    - nonenrollers in higher education;
    - prompt (within 6 months or a year of high school graduation) postsecondary education enrollers;
    - delayed postsecondary education enrollers; and
    - postsecondary leavers (versus persisters) and returnees.

- The second follow-up also captures three distinct (and sometimes alternating or combined) transitions:
  - transition from high school to the work force;
  - transition from high school to postsecondary education; and
  - transition to adult roles.

The second follow-up in the spring of 2006 employed a web-based self-administered instrument with CATI and CAPI data collection for nonresponse follow-up. The focus of the interview was on transition to postsecondary education and labor force participation. Out of a sample of about 15,900 cases, about 14,200 sample members completed interviews.

The ELS:2002 second follow-up provides data to map and understand a key transition: the path from high school to adulthood, as seen in the labor market, postsecondary education, family formation, and civic engagement. The second follow-up collected information to enable researchers and policymakers to better understand issues of postsecondary educational access and choice. Thus, a major focus of the second follow-up was the postsecondary decision-making process as reflected in applications to college and initial postsecondary enrollment histories. ELS:2002, unlike studies that sample only postsecondary students, is uniquely positioned to address these issues because it tracks respondents who attended postsecondary institutions before they enrolled. Additionally, it follows students who did not attend college and thus provides information on reasons students did not attend. The second follow-up also provides information about high school completion (for students who dropped out or were held back), as well as information about the status of dropouts and students who have obtained an alternative credential, such as the GED. For non–college-bound students, the second follow-up mapped the transition into the labor market. In addition to its focus on postsecondary (or sometimes secondary) education and work experiences, the second follow-up survey also obtained information about family formation. Indeed, workforce data are collected on everyone, including those who are also enrolled in postsecondary education, thus supplying workforce data on all cohort members who are employed full- or part-time. For further information about ELS:2002 base year through second follow-up, see Ingels et al. 2007

Third Follow-up Field Test (2011) and Main Study (2012)

Field Test. The purpose of the 2011 field test was to test the proposed third follow-up survey instrument and basic methodologies and procedures. Evaluating the questionnaire includes obtaining timings, both overall and for each item. It also involves examination of item distributional properties, to ensure that there is meaningful variance, captured appropriately by the response categories, and that scales are reliable, and comprise the optimal items needed to measure the underlying construct. Other forms that are evaluated include respondent letters and brochures—testing the entire apparatus of data collection is a major field test objective. The data collection challenges of the third follow-up round are substantial, and therefore the evaluation of data collection methods and procedures is especially critical. Although address updates were

conducted, a formidable locating task remains, given that the age of the cohort marks a period of high mobility, and given that the previous interview took place 6 years earlier. Finally, the 2011 field test also afforded a limited experiment into case response propensity modeling, monitoring of data collection, and case prioritization with a focus on minimizing nonresponse bias and maximizing the stability of estimates. (This is an alternative strategy to the tradition of monitoring through response rates and assuming that a high response rate will necessarily result in low bias.) Protocols and procedures for the ELS:2002/12 postsecondary transcripts study will be evaluated in a separate, pilot study. The results of the 2011 field test, and recommendations for the main study, are set out in the various chapters and appendixes that follow.

Main Study. The third follow-up sample design is straightforward: all members of the second follow-up sample, regardless of their 2006 participation status, will be pursued in the third follow-up, with the exception of the deceased. In addition, cases comprising study withdrawals who have asked to not be recontacted and those who did not respond in any round, although technically eligible, will not be fielded. Some fielded cases will be found to be out of scope for the 2012 round—interviews will not be attempted if the sample member is incarcerated or institutionalized or out of the country.

An approximately 35-minute web-based electronic questionnaire will be employed, with the same questionnaire also available for CATI and field interviews. The focus of the questionnaire records final outcomes across three topical areas:

- employment and career outcomes;
- education outcomes; and
- other outcomes reflective of attaining to adult status such as family formation, life and career values, and civic engagement.

Products of the full-scale study will include a volume of Data File Documentation (DFD) that covers instrumentation; sample design; data collection results and methods; data preparation and planning; weighting, imputation, and design effects; and data file contents. Appendixes to the DFD will include a facsimile of the electronic questionnaire and a hard-copy codebook with response frequencies. There will also be a “First Look” report that sets out some basic descriptive findings, and public and restricted use data sets, with suitable analysis systems.

This page left intentionally blank.

Chapter 2.
Field Test Survey Design and Preparation

Chapter 2 reports on key elements of survey design and preparation for the 2011 field test: instrument development (including its goals and constraints and the instrument development process), and pre–data collection activities (chiefly recruitment and training of data collectors).

2.1 ELS:2002 Third Follow-up Data Collection Instrumentation

2.1.1 Instrument Development Goals and Constraints

The Education Longitudinal Study of 2002 (ELS:2002) research agenda was set in the 2002 base year with the later rounds fully in mind. In the very beginning, key information about the hypothesized antecedents of later outcomes was identified and collected. ELS:2002 was intended to constitute a general purpose data set that would allow researchers from a variety of disciplines to examine changes in young people's lives as they transitioned to adulthood, and to examine their connections with communities, schools, teachers, families, parents, and friends. Research issues include the following:

- academic, social, and interpersonal growth;
- transitions from high school to postsecondary education and from school to work;
- the characteristics of schools and colleges and their impact on student outcomes;
- civic engagement and the transition to adulthood;
- family formation, including marriage, and how prior experiences in and out of school correlate with these decisions; and
- the contexts of education, including how minority and at-risk status is associated with education and labor market outcomes.

There is much competition of worthy constructs and items for the limited space on the third follow-up questionnaire. The need to prioritize is driven by important constraints on what the third follow-up of ELS:2002 can accomplish. First, it is the final round of the study, hence time to collect outcomes, not predictors. Second, the cohort has followed highly varied pathways into postsecondary education, the work force, and other adult roles—the questionnaire must accommodate all, from high school dropouts to PhD candidates, from workers to stay-at-home parents. Third, length of interview—although operationally quite reasonable—is a content constraint: the average interview will be around 35 minutes. The time span to be covered—6 years—marks a further constraint. A sub-problem of the 6-year span to be covered is the need to retrieve information about the past without risking recall bias, a pervasive danger but especially

for all but the most salient experience and attitude questions. The third follow-up interview will be supplemented with linkages to various administrative records, including postsecondary educational transcripts. It is therefore critical that the questionnaire identify the names, locations, and Integrated Postsecondary Education Data System codes for all postsecondary institutions attended by sample members since high school, but not necessary to collect information on the questionnaire that can be obtained from transcripts.

As a general principle, the 2012 content should focus on the unique features and special strengths of the study by gathering outcomes for the areas in which it has already invested by collecting critical antecedents. Information that uniquely fulfills the ELS:2002 design should be collected in preference to duplicating the effort of other surveys that may be better suited than ELS:2002 to tackle a particular issue. For example, the Department of Labor's National Longitudinal Surveys, youth cohorts (NLSY) is better equipped to collect complete labor market event histories, because the NLSY interview is longer and the periodicity is one of far more frequent interviews. The Beginning Postsecondary Students (BPS) study is being specifically redesigned to optimally capture persistence in postsecondary education and in field of study. BPS is better able to accommodate the fact of and to study nontraditional late entrants into postsecondary education, while ELS:2002 is better able than BPS to examine issues of choice because its cohort includes, prospectively, both those who do and who do not enter postsecondary education. Indeed, postsecondary educational choice was the keynote of the ELS:2002 second follow-up round in 2006, when the modal cohort age was 20, just as educational attainment and early career experience will be the keynotes of the 2012 third follow-up, when the cohort modal age is 26. Items that have been asked in prior NCES studies, such as the National Education Longitudinal Study of 1988, have important extra value if repeated in ELS:2002 in that they support cross-cohort analysis. However, these potential comparison items should not be permitted to supplant more useful new constructs and items for the current study, if the new constructs and items work longitudinally and promise fresh insights.

2.1.2 Instrument Development Process

The field test instrument development process can be concisely described. First, a literature review was conducted. Second, questionnaire drafts were circulated across programs at the National Center for Education Statistics (NCES) and between RTI International and NCES, and revisions made on the basis of the resulting comments. Third, new topics and constructs, and prioritizations of these and past constructs, were recommended in commissioned papers from three ELS:2002 Technical Review Panel members, representing three distinct academic disciplines—Professors Robert Lent (psychology), Randall Olsen (labor economics), and Michael Shanahan (sociology). These memoranda were shared with the other panelists.

Fourth, certain survey items that were new and of critical importance were tested in two series of cognitive interviews (a report on cognitive research for ELS:2002 may be found in appendix D). The first series of cognitive interviews (pre-field test) focused on social cognitive

career theory items that were developed by Professor Lent; the second series (pre–full-scale study) focused on potential new items concerning financial aid and economic literacy.

Fifth, before and after the field test, draft instruments were reviewed by the ELS:2002/12 Technical Review Panel, a specially appointed independent group of substantive, methodological, and technical experts. A summary of the two panel meetings (held in September 2010 and November 2011) is found in appendix A. Sixth, justifications were written for questionnaire items for Office of Management and Budget (OMB) review.

Seventh, actual development of the questionnaires, including specification, routing, programming, and testing, took place within an RTI proprietary system, the Hatteras Survey Engine and Survey editor. (A description of Hatteras appears in chapter 5.) Eighth, a field test was conducted to test questionnaire and assessment items; that field test is documented in this report, which includes recommendations for the main study instruments and procedures.

A hardcopy version of the electronic field test questionnaire is found in appendix C. The quality of the questionnaire data is addressed in chapter 4.

2.2 Pre-Data Collection Activities

The following sections describe the interviewer training and sample locating activities for the ELS:2002/12 third follow-up field test completed in preparation for data collection.

2.2.1 Training of Data Collectors

The data collection staff included quality control supervisors (QCSs), quality experts (QEs), telephone interviewers (TIs), and intensive-tracing staff. Prior to beginning work on ELS:2002/12, all data collection staff completed a comprehensive training program. Topics covered in the training program included the following:

- a review of confidentiality requirements;
- an overview of the study;
- frequently asked questions (FAQs);
- administrative procedures for case management; and
- hands-on practice.

The training program was designed to maximize active participation of the trainees. Thirteen data collection staff, including interviewers, QCSs, and QEs, were trained. Help desk support staff were trained on July 10, 2011, and interviewers were trained on help desk support and interviewing on July 27–29, 2011. The specific roles and duties of data collection staff are summarized in the following subsections, along with a description of the training program.

Quality Control Supervisors and Quality Experts. QCSs provided support and guidance for the telephone interviewers and helped troubleshoot problems, and QEs monitored interviewer production. They attended ELS:2002/12 project supervisor training and also

participated in telephone interviewer project training. The project supervisor training included an overview of the study, active listening techniques expected of interviewing staff, problem resolution, and other specific project procedures and protocols.

Telephone Interviewers. As the primary point of contact with sample members, TIs were responsible for gaining cooperation from and conducting interviews with sample members, avoiding interview refusals, and addressing the concerns of reluctant sample members. Because of this integral role in the study, TIs received a project manual and 12 hours of training that included

- an overview of ELS:2002/12;
- an in-depth review of the questionnaire;
- hands-on practice administering the telephone interview;
- review of appropriate conversational interviewing techniques; and
- review of proper methods in handling inbound calls.

To conclude the training and verify that the material had been learned, all telephone interviewers were certified by successfully conducting mock telephone interviews and by providing satisfactory responses to the study's FAQs.

In addition to the training described above, QCSs and TIs were also cross-trained as help desk agents who assisted any sample members who had questions or problems while completing web interviews or called in to complete the telephone interview.

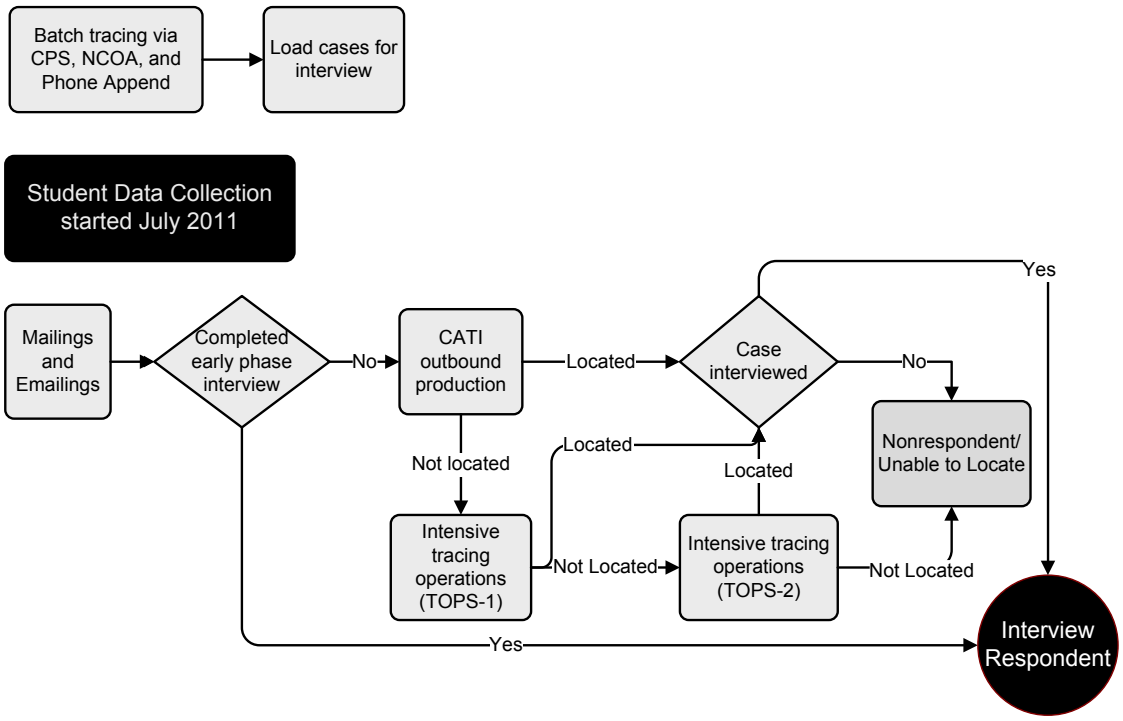
Tracing Staff. Tracing staff (tracers) used intensive measures, described in section 2.2.2, to locate sample members who lacked good telephone contact information. Tracers attended a comprehensive 16-hour training session led by RTI tracing managers and covered all tracing procedures. ELS:2002/12 tracers were cross-trained as TIs on the same project.

Additionally, weekly quality circle meetings were routinely conducted as an extension of the training program for continual quality improvement. Quality circle meetings are discussed in greater detail in section 3.1.

2.2.2 Locating and Tracing Activities

Several locating methods were used to find and collect up-to-date contact information for the ELS:2002/12 sample (figure 2). Batch searches of national databases and address update mailings to sample members were conducted prior to the start of data collection. Follow-up locating methods were employed for those sample members not found after the start of data collection, including computer-assisted telephone interview (CATI) locating and intensive tracing.

Figure 2. ELS:2002/12 tracing procedures: 2011



NOTE: CATI = computer-assisted telephone interviewing. CPS = Central Processing System. NCOA = National Change of Address.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Second Follow-up (ELS:2002/12).

Batch Tracing. Batch database searches were conducted on all sample members to update contact information in preparation for mailing activities. These searches used the U.S. Department of Education's Central Processing System and the U.S. Postal Service (USPS) National Change of Address databases. Then, just prior to the start of outbound telephone interviewing, all sample members were sent to Phone Append, which searches more than 400 million landline, voice over Internet protocol, and wireless numbers in the United States, Puerto Rico, and Canada. All information obtained from these sources was compared with the information previously available from the ELS:2002/12 second follow-up locator database to identify any new contact information.

Mailings. An initial letter announcing the data collection was mailed on July 12, 2011, by USPS first-class mail. The mailing to all sample members included the following:

- a letter, signed by both the project director and the NCES project officer, that announced the start of data collection;
- information about the incentive;
- a link to the study website;
- login credentials for accessing the web interview;
- the ELS e-mail address and toll-free help desk number; and
- a brochure about ELS:2002/12.

In addition to the letter, an e-mail was sent to sample members 2 days later, on July 14, 2011. Both the letter and the e-mail encouraged sample members to complete the survey online during the early response period. Additional mailings during this early period included a reminder postcard and letter and additional e-mail reminders to encourage early interview response. Once outbound telephone interview efforts began, periodic reminder mailings and e-mails were sent to sample members throughout the course of data collection. Copies of letters and brochures can be found in appendix B.

CATI Locating and Pre-intensive Tracing. Once outbound telephone interviewing began, telephone interviewers conducted limited tracing and locating activities, as needed. The telephone number believed to be the best known number for contacting the sample member was attempted first. If the sample member could not be reached at that number after several attempts, any other numbers associated with the sample member, including parent and other contacts, were called. The interviewers attempted to gather locating information for the sample member from the contact who answered the call. If the sample member could not be located, the case was sent for intensive interactive tracing by RTI's Tracing Operations (TOPS).

Intensive Tracing. Cases that could not be located through batch tracing or CATI locating efforts underwent intensive tracing by RTI's TOPS. Using a number of public domain and proprietary databases, TOPS uses a two-tiered strategy. The first tier (TOPS-1) used Social Security numbers (SSNs) to search for sample members in consumer databases such as FastData's SSN search and Experian which contains current address and telephone listings for the majority of consumers with credit histories. If a search generated a new telephone number for the sample member, tracers attempted to confirm the information by speaking with the sample member or with someone else who could confirm the information. If the number was confirmed, the case was sent back to CATI for telephone interviewing. This first level of effort minimized the time that cases were in tracing and unavailable for CATI efforts.

If cases were not located (or locating information not confirmed) at the end of TOPS-1, the cases underwent a more intensive level of tracing in TOPS-2, which included calls to other possible sources of information, such as directory assistance, alumni offices, and contacts with neighbors or landlords. Whenever any of these sources provided information that indicated that a sample member was not available for the study (e.g., deceased, incarcerated, or out of the country), no further contact efforts were made.

Chapter 3.
Data Collection Procedures and Results

Data collection procedures and outcomes for the field test will be covered in this chapter. The procedures section describes the phases of data collection and modes of interview administration. The outcomes section includes the locating rate and response rates by data collection phase and mode as well as high school dropout status and second follow-up response status. The response propensity experiment procedures and results are described in the final section.

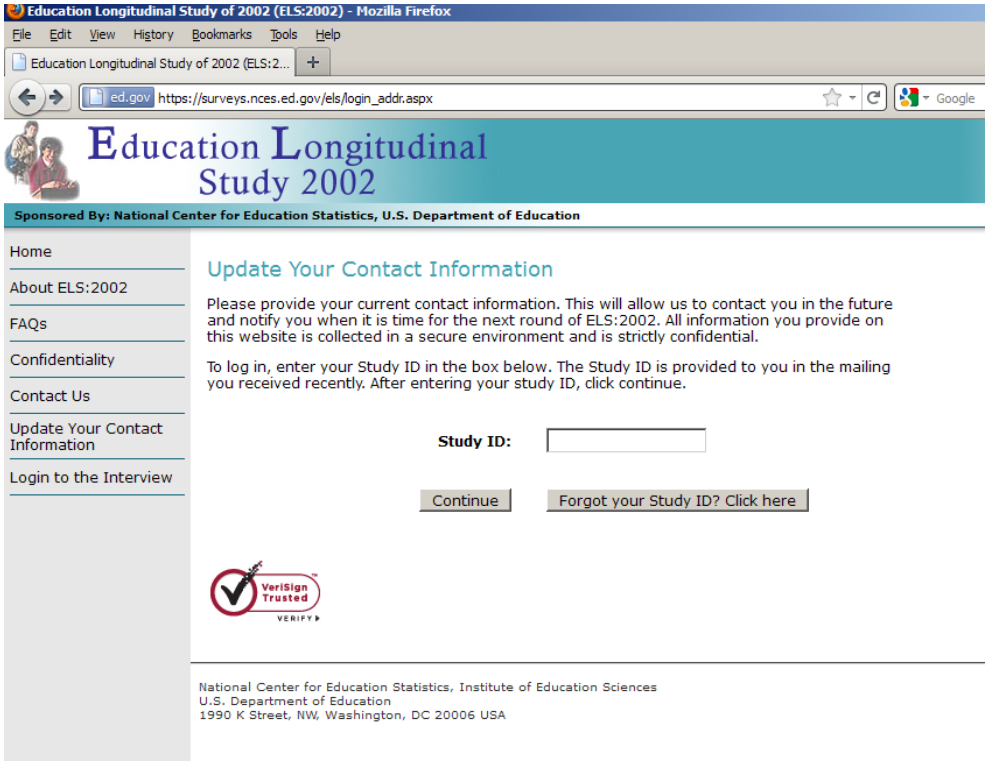
3.1 Web/CATI Data Collection Procedures

3.1.1 Study Website and Help Desk

Sample members were provided with a link to the Education Longitudinal Study of 2002 third follow-up (ELS:2002/12) website prior to the start of data collection. The website provided general information about the study, including the study sponsor and contractor, how the data are used, answers to frequently asked questions (FAQs), confidentiality assurances, and selected findings from earlier rounds of ELS:2002. The website also provided contact information for the study help desk and project staff at RTI, as well as a link to the National Center for Education Statistics (NCES) website. Sample members were able to log in to the secure website to provide updated contact information and complete the sample member interview once it became available.

Designed according to NCES web policies, the study website used a three-tier security approach to protect all data collected. The first tier of security included secure logins, with a unique study ID and strong password provided to sample members. The second tier of security protected any data entered on the website with secure socket layer technology, allowing only encrypted data to be transmitted over the Internet. The third tier of security stored any collected data in a secured SQL Server database located on a server machine that was physically separate from the web server. Figure 3 shows the home page for the study website.

Figure 3. ELS:2002/12 website home page: 2011



Sample members were provided with a toll-free telephone number, which was answered by help desk agents. Help desk staff were available to sample members who had questions or technical issues related to completion of the web interview. For each call received, staff confirmed contact information for the sample member and recorded a description of the problem and resolution. If technical difficulties prevented sample members from completing the web interview, help desk staff were able to complete a telephone interview. Two common types of help desk incidents were requests for login credentials and requests to complete the interview over the telephone. To minimize the need for help desk assistance, a "Forgot Password?" link was included on the study website and the need to disable pop-up blockers to launch the survey was eliminated.

3.1.2 Interviewing

Data collection for the ELS:2002/12 field test interview consisted of three phases:

1. *Phase 1: Early Response*. This phase began with the start of data collection in July 2011 and lasted approximately 3 weeks until August 1, 2011. Sample members were encouraged to complete the survey over the Web during this phase. The telephone interview was available to sample members who contacted the help desk, but no outbound telephone calls were made. Sample members who completed the interview were eligible to receive an incentive of $25 or $45 based on their response propensity experiment group assignment (see section 3.2 for more information on the

experiment). The sample members were divided into two groups, high propensity and low propensity. Sample members in the high-propensity group were offered $25 while half of the low-propensity cases were offered $45 (experimental group) and half were offered $25 (control group).

2. *Phase 2: CATI Production.* The computer-assisted telephone interviewing (CATI) production phase began on August 1, 2011. During this phase, interviewers called to encourage sample members to complete the interview by telephone or on the Web. Sample members who completed the interview during the production phase were eligible to receive the same incentives as during the early response phase.

3. *Phase 3: Increased Incentive*. The increased incentive phase began on September 13, 2011. During this phase, the incentive values were increased to $35 for high-propensity cases and low-propensity control cases and $55 for low-propensity experimental cases. Sample members were alerted with a letter and e-mail, and by interviewers if successful contact was made.

Sample members could complete the interview on the Web or by telephone throughout the entire data collection period. For the telephone interviewer, the interview screens were identical to those in the web interviews completed by respondents, except that instructions on how to administer each question were visible at the top of each screen for telephone interviews. Following are details of the administration of the interview through the various modes.

Web Interviews. Sample members were informed of the web interview in all project communications including mailings, e-mails, and telephone contacts. During the early response period (the first 3 weeks of data collection), only web interviews were completed unless sample members initiated a telephone interview by calling the help desk or sending an e-mail asking to be called. Reminder mailings and e-mails were sent throughout all phases of data collection to encourage sample members to complete the interview online. The website was accessible 24 hours a day, 7 days a week, throughout the data collection period, providing sample members with the option to complete the interview online at any time.

Telephone Interviews. Outbound telephone follow-up began on August 1, 2011, after the 3-week early response period ended. Telephone interviewers attempted to locate, gain cooperation from, and interview sample members who had not yet completed the interview. Interviewers encouraged sample members to complete the interview by telephone; however, the web survey remained available throughout data collection for sample members who preferred that option. Sample members who did express a preference to complete a web interview were called back 5 days later for follow-up if the interview had not yet been completed.

The CATI Case Management System (CATI-CMS) included an automated call scheduler that assigned cases to interviewers by case priority, time of day, day of week, existence of previously scheduled appointments, and type of case. Case assignment was designed to maximize the likelihood of contacting and interviewing sample members, and cases were assigned to various queues accordingly. For example, the CATI-CMS included queues for new cases that had not been called, Spanish-language cases, initial refusals, and various appointment

queues. In addition, available telephone numbers for each case were automatically prioritized for the interviewers. As new telephone numbers were added—as a result of CATI tracing, other tracing efforts, and information from other sources such as respondent e-mails or help desk call-ins—available telephone numbers were reprioritized based on the new information.

Some cases required special treatment. To gain cooperation from those sample members who initially refused to participate (and from contacts such as parents and roommates who acted as *gatekeepers* to the sample member), interviewers were trained in refusal conversion techniques. For example, a telephone interviewer will recontact the person and acknowledge the issue or concern the person conveys and listen carefully. He or she will then speak to the specific concern and also explain the importance of the study and the importance that individuals similar to the sample member are represented in the data.

Help Desk Operations. Throughout ELS:2002/12 field test data collection, telephone interviewers and quality control supervisors served as help desk agents for the toll-free help line. They were available to assist sample members who had questions or problems accessing and completing the web interview or who called in to complete the telephone interview. If technical difficulties prevented sample members from completing the web interview, sample members were encouraged to complete a telephone interview.

Quality Circle Meetings. Weekly quality circle meetings were conducted to serve as an essential feedback loop for ensuring that project staff and call center staff were communicating on a regular basis about the goals of the study and addressing challenges encountered along the way. These meetings provided a forum for discussing elements of the instrument design and interview cooperation tactics, motivating the group toward the goals of the study, and obtaining feedback on data collection issues. Weekly quality circle meetings for telephone staff were held at the call center. Issues discussed at these meetings were documented in meeting notes and stored where staff could access them as needed. The interviewers were informed of counts of interview completions to date, general data collection issues and issues specific to the survey instrument, and project staff responses to questions from interviewers.

Throughout the study, a variety of issues were addressed at the quality circle meetings that reinforced specific content from training and contributed to prompt problem solving. Some of the issues covered in these meetings included the following:

- clarification of questions and item responses and reinforcement of positive interviewing techniques;
- methods of gaining cooperation from sample members and gatekeepers (e.g., spouses, parents, and roommates);
- problem sheets submitted during contacting and interviewing;
- the importance of providing and reviewing detailed case comments;
- data security protocols; and
- study progress and general morale boosting.

Telephone interviewers and their supervisors were given the opportunity to ask questions in quality circle meetings, and as needs were identified, additional training topics were highlighted and addressed in subsequent meetings.

3.2 Interview Data Collection Outcomes

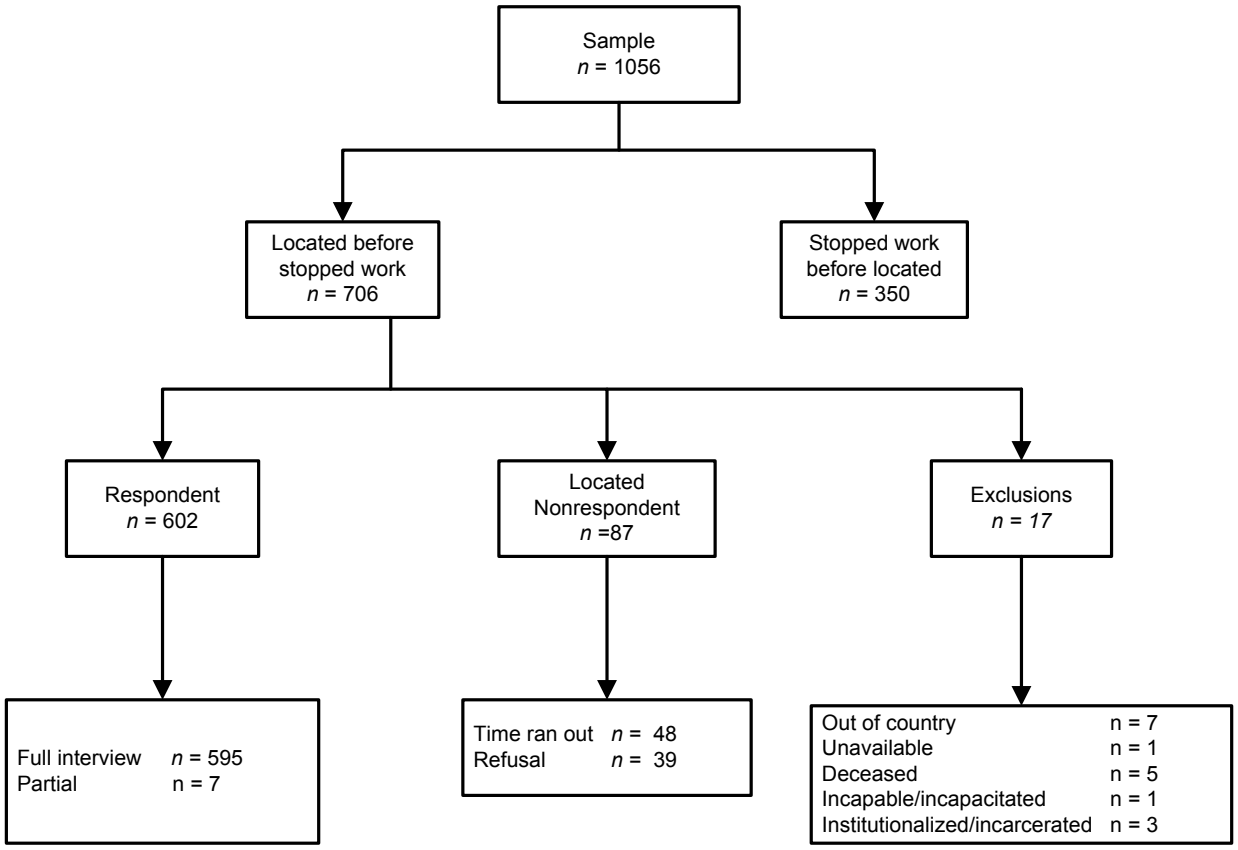
The 2011 field test inherited a substantially larger sample than was needed or fully used. In the in-school rounds (base year when the participants were sophomores and first follow-up when the majority of participants were seniors), a field test sample of at least 1,000 was required to perform the psychometric analyses required for direct cognitive assessment development. The pool of respondents needed for questionnaire analysis, however, is about half that amount. This fact argued for cessation of data collection after reaching a reasonable threshold (500) for questionnaire item analysis. In turn, response rates reflect the fact that not all cases were pursued, and are not indicative of the response rates expected in the full-scale study.

ELS:2002/12 field test interviews were administered between July 13, 2011, and September 30, 2011. Of the 1,056 field test sample members, 1,039 were eligible for the study after removing those who were ineligible for reasons such as being institutionalized, incarcerated, or out of the country. Of these eligible members, 602 sample members (58 percent) had completed a full interview or a partial interview[7] when data collection was stopped. The overall locating and interviewing results for the ELS:2002/12 field test data collection effort, including sample members who were located but later excluded, are presented in figure 4. The figure displays counts at the time data collection was stopped because the yield goal was achieved and surpassed. A sample member was considered located if, as of the day we stopped work, an interview was completed, the sample member was successfully contacted despite not completing the interview, or the contact information in the study database had not been proven out of date or incorrect during contact attempts.

In the full-scale study, the goal will be to achieve an overall response rate of 90 percent. A direct comparison of the response rate achieved in the field test with the needed response rate in the full-scale study is not appropriate for a number of reasons. First, field test data collection stopped soon after surpassing the target interview yield; the full-scale data collection period will be more than 6 months in duration as compared to 2.5 months in the field test. Additional time to locate sample members and convert refusals will increase the response rate. In addition, the field test sample has a greater proportion of second follow-up nonrespondents than the full-scale sample. As will be illustrated later in this section, second follow-up nonrespondents, as well as high school dropouts, were particularly challenging to locate.

[7] A partial interview is defined as any case where the respondent began the interview, completed it through the high school completion section, but broke off and did not return to complete the interview.

Figure 4. Overall locating and interviewing rates at the time work stopped



Locating Results and Interview Response Rates. Overall snapshot locating rates and response rates for the ELS:2002/12 interview were determined at the point that data collection ended early. Thus, the rates are only intended to serve as a comparison between groups and not indicative of final rates expected from a data collection that runs its full course. The snapshot locating and response rates varied by high school dropout status and prior-round response status, as shown in table 1. As alluded to above, the locating rate was defined as the percentage of cases in which an interview was completed, the sample member was successfully contacted, or the contact information in the study database was not proven out of date or incorrect during contact attempts. The snapshot locate rate was significantly higher for those who never dropped out compared to those who ever dropped out of high school ($t = 2.83$, $p < .05$). Regarding response rate, of field test sample members who ever dropped out, the snapshot response rate was significantly lower compared to those who never dropped out ($t = 3.49$, $p < .05$).

Table 1. Student locating snapshot, by dropout status and second follow-up response status: 2011

ELS:2002 dropout status and second follow-up response status	Total eligible	Located snapshot		Responding sample members		
		Number	Percent of total	Number	Percent of located	Percent of total
Total	1,039	689	66.3	602	87.4	57.9
Dropout status						
Ever dropped out	95	50	52.6	39	78.0	41.1
Not ever dropped out	944	639	67.7	563	88.1	59.6
Second follow-up response status						
ELS:2002/06 respondent	771	560	72.6	502	89.6	65.1
ELS:2002/06 nonrespondent	268	129	48.1	100	77.5	37.3

NOTE: Numbers may not sum to total because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

In addition to differences by high school dropout status, snapshot locate and response rates differed by prior-round (ELS:2002/06) response. The snapshot locate rate was significantly higher for those who responded to ELS:2002/06 (second follow-up) compared to those who did not respond to the second follow-up ($t = 7.10$, $p < .05$). Interview snapshot response rates also differed with a significantly higher percentage of second follow-up respondents completing the third follow-up field test interview in contrast to second follow-up nonrespondents ($t = 8.14$, $p < .05$).

A concerted effort must be made in the full-scale study to convince parents of the value of the study so they cooperate and share information with their children. It was learned that in some cases parents continue to act as gatekeepers for the sample members even though the sample members are several years removed from being minors (15 percent of refusals in the field test were by someone other than the sample member). The full-scale study procedures include dual mailings to both the sample members and their parents. These procedures will ensure at both the panel maintenance and data collection stages that parents are aware of the procedures for the third follow-up. Direct mail and e-mail contacts with parents will also allow parents to provide updated contact information for their young adults. Also, because some phone numbers available for sample members are numbers for their parents' homes, CATI procedures will be in place to guide interviewers on how to appropriately ask for and record new contact information for sample members from parents. Successful contacts with parents will be an important part of interviewer training for the full-scale study.

Locate Rates by Source of Batch Update. Prior to the ELS:2002/12 field test data collection, batch matching was conducted for the 1,056 sample members. This effort confirmed contact information or provided new contact information for several hundred field test cases, as shown in table 2. The highest proportion of matches was obtained through FirstData (National Change of Address and Phone Append), 62 percent of cases sent in 2010 and 58 percent in 2011. The smallest proportion, 20 percent of cases sent in 2010 and 19 percent in 2011, were matched

through the Central Processing System (CPS). Because CPS relies on postsecondary student loan application data, it is unsurprising that CPS matched the fewest records because only 65 percent of ELS:2002/06 respondents reported attending postsecondary education courses in the previous interview, and it had also been 8 years since the majority of sample members completed high school so the numbers attending postsecondary institutions were likely to be low. Also, not all postsecondary students fill out the application for federal student aid and those students would not appear in the CPS data.

Table 2. Batch processing confirmation/update rates, by tracing source for 2010 and 2011: 2011

	2010 batch tracing			2011 batch tracing		
Method of tracing	Number of cases sent	Number of confirmed/ updated	Percent	Number of cases sent	Number of confirmed/ updated	Percent
Central Processing System	984	194	19.7	543	103	19.0
FirstData (NCOA, Phone Append)	1,049	652	62.2	986	570	57.8
Accurint	348	326	93.7	176	141	80.1

NOTE: Batch tracing was conducted prior to data collection with the sample of 1,056 cases. NCOA = National Change of Address.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Address Update Mailing Results. The last panel maintenance address update mailing for the cohort, in which sample members and one parent are asked to confirm or update contact information, was conducted in fall/winter of 2010. Address updates or confirmations were received from 230 field test sample members (22 percent) in response to the request, which was sent via mail and e-mail. The response was double the response seen in the 2007 and 2008 address update efforts (11 percent and 10 percent, respectively). An incentive experiment was conducted during the 2010 panel maintenance effort to assess whether offering a $10 incentive would increase the response rate for an address update mailing. The sample was split in half. One half was offered $10 if the sample member or a parent confirmed or updated contact information, while the other half was offered no incentive. The experiment results showed that the $10 group had a significantly higher address update response rate (25 percent) than the $0 group (20 percent, one-tailed $t = 1.90$, $p < .05$), as shown in Table 3. Based on the positive experiment results, ELS:2002/12 plans include offering a $10 incentive to full-scale sample members if they or a parent updates or confirms contact information when the fall 2011/winter 2012 address update mailing is conducted.

Table 3. Address update participation by incentive group: 2011

Type of incentive	Eligible sample[1]	Participated	
		Number	Percent
$0	495	99	20.0
$10	489	124	25.4

[1] Eligible sample includes sample members for whom an address was known and the mailing was not returned as undeliverable.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2002 Education Longitudinal Study (ELS:2002/12) Third Follow-up Field Test.

Interview Completion by Address Update Participation. Sample members who responded to the panel maintenance address update request responded to the third follow-up interview at a higher rate than those who did not respond to the request (one-tailed $t = 18.28$, $p < .05$). Panel maintenance address update request nonrespondents responded to the interview at a higher rate if offered $10 during panel maintenance efforts (one-tailed $t = 1.84$, $p < .05$). In contrast, panel maintenance respondents showed only a small difference in interview response rate if offered (and paid) $10 during panel maintenance. The observed difference was not statistically significant. Table 4 contains the third follow-up field test interview completion rates by panel maintenance address update response.

Table 4. Interview completion rate, by address update and incentive: 2011

	Number of eligible cases	Interviewed	
		Number	Percent
Total	1,039	595	57.3
Panel maintenance respondent	228	211	92.5
Offered address update incentive	126	118	93.7
Not offered incentive	102	93	91.2
Panel maintenance nonrespondent	811	384	47.3
Offered address update incentive	418	214	51.2
Not offered incentive	393	170	43.3

NOTE: This table includes the completed interviews (595), not the additional partial interviews (7).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Sample Members Requiring Intensive Tracing. Overall, about 160 (16 percent) of the 1,039 eligible field test sample members required intensive tracing (table 5) before data collection was closed early. Thirty-three percent of sample members who ever dropped out of high school required intensive tracing, compared with 14 percent of sample members who never dropped out. Thirty-six percent of second follow-up nonrespondents required intensive tracing, compared with 9 percent of second follow-up respondents. Of the cases located through intensive tracing before data collection ended early, about 73 percent completed the ELS:2002/12 interview.

Table 5. Sample members requiring intensive tracing procedures, by dropout status and second follow-up response status: 2011

Dropout status and second follow-up response status	Total	Cases requiring intensive tracing	
		Number	Percent
Total	1,039	163	15.7
Dropout status			
Ever dropped out	95	31	32.6
Never dropped out	944	132	14.0
Second follow-up response status			
ELS:2002/06 respondent	771	66	8.6
ELS:2002/06 nonrespondent	268	97	36.2

NOTE: Numbers may not sum to total because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Interview Outcomes by Mode. ELS:2002/12 interviews were completed on the Web or over the telephone. Table 6 shows that 35 percent of interviews were completed on the Web without telephone contacts, 48 percent of interviews were completed on the Web after being contacted by a telephone interviewer, and 17 percent were completed on the telephone. Summed, it was 83 percent of interviews completed on the Web. This is in line with the expected increased Internet use over time and the greater proportion of web interviews expected closer to the beginning of the data collection period. The prior-round field test produced 37 percent of the completed interviews on the Web, although it is important to keep in mind that the third follow-up field test took place over a shorter time period. A longer time period for data collection would likely have resulted in a greater proportion of telephone interviews than seen in the current field test.

Table 6. Interview completion rate, by address update and incentive: 2011

	Interviewed	
	Number	Percent
Web without telephone contacts	212	35.2
Web with telephone contacts	286	47.5
Telephone interview	104	17.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Response by Phase of Data Collection. Interview response, by phase of data collection, is shown in table 7. Thirty-six percent of completed ELS:2002/12 interviews were completed during the early response phase. Fifty-one percent of interviews were completed during phase 2, which added telephone prompting and interviewing. Thirteen percent were completed during phase 3, when the incentive amount was increased. Section 3.3 contains details of the phases of data collection as they relate to the response propensity experiment.

Table 7. Distribution of completed interviews, by data collection phase: 2011

	Interviewed	
	Number	Percent
Eligible	1,039	†
Respondent	602	57.9
Early response phase[1]	217	36.0
Computer-assisted telephone interview phase	307	51.0
Increased incentive phase[2]	78	12.9

† Not applicable.
[1] Of the 217 respondents, 212 were complete and 5 were partially complete.
[2] Of the 78 respondents, 76 were complete and 2 were partially complete.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Each phase of data collection can be evaluated individually. The early response phase of data collection yielded 217 cases that were either complete or partially complete. Of the 217 cases, 212 were fully complete, translating to a 20 percent response rate (212 completed out of 1,039 eligible) (table 8). The next phase of data collection yielded a 37 percent response rate (307 completed out of 827 eligible for the second phase). The third phase yielded 78 cases that were either complete or partially complete. Of the 78 cases, 76 were fully complete, translating to a 15 percent response rate (78 completed out of 520 eligible by the third phase). As mentioned previously, the target yield was met prior to the third phase, but it was conducted briefly to test procedures and determine whether additional yield would result from those procedures.

Table 8. Number of cases and percentage of completed interviews by data collection phase: 2011

Data collection phase	Number of cases eligible for phase	Completed interviews	
		Number	Percent of cases
Early response phase	1,039	212	20.4
Computer-assisted telephone interview phase	827	307	37.1
Increased incentive phase	520	76	14.6

NOTE: Partial interviews were not included because partially completed interviews could be resumed by sample members through the end of data collection. A partial interview is defined as any case where the respondent began the interview, finished at least through the high school completion section, but broke off and did not return to complete the interview.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Telephone Interviewer Hours and Call Counts. During the course of ELS:2002/12 data collection 100 telephone interviews were completed. On average, telephone interviews that were completed in a single session took 41.3 minutes to administer. Most of the telephone interview hours not spent administering an interview were spent on case management activities such as locating and contacting sample members, prompting sample members to complete interviews, reviewing call history, scheduling appointments for callbacks, and entering detailed comments and suggestions to assist with reaching and interviewing sample members. Some of the time was also spent responding to incoming help desk calls.

On average, 13 calls were made per ELS:2002/12 field test sample member during the data collection period.[8] Of the sample, cases that completed the current-round (the third follow-up) interview required an average of 7 calls, while current-round nonrespondents received an average of 21 calls during the interviewing period. There were no significant differences in call counts between respondents who completed interviews over the telephone and respondents who completed interviews over the Web but with phone prompting. The average number of telephone calls is shown in table 9.

Table 9. Number and average of calls, by dropout status and prior- and current-round response status and mode of interview: 2011

Response status and mode	Number of cases	Number of calls	Average number of calls
Total	1,039	13,527	13.0
Dropout status			
Ever dropped out	95	1,361	14.3
Never dropped out	944	12,166	12.9
Second follow-up response status			
ELS:2002/06 respondent	771	9,708	12.6
ELS:2002/06 nonrespondent	268	3,819	14.3
Current-round response status			
Respondent	602	4,233	7.0
Web interviews	498	3,168	6.4
Excluding early response	287	3,168	11.0
Telephone interviews	104	1,065	10.2
Nonrespondent and exclusions	454	9,367	20.6

NOTE: Numbers may not sum to total because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Averting and Converting Refusals. Refusal aversion and conversion techniques were integrated into telephone interviewer training and were reinforced throughout data collection in quality circle meetings. Interviewers were encouraged to share their experiences gaining sample member cooperation and seek guidance from the group. Sample members who refused to complete the interview were placed in a separate queue and worked by a subset of interviewers selected for additional refusal conversion training. Overall, 10 percent of eligible cases ever refused. Of these refusals, 21 percent subsequently completed the interview (table 10) after refusal conversion attempts were made. In the previous round (second follow-up), 13 percent of eligible cases ever refused and, of these, 8 percent subsequently completed the interview.

[8] This includes sample members who required no call attempts.

Table 10. Refusal and refusal conversion rates, by dropout status and prior-round response status: 2011

		Ever refused interview		Interviewed, given refusal		
	Total eligible	Number	Percent of total	Number	Percent of refused	Percent of total eligible
Total	1,039	103	9.9	22	21.4	2.1
Dropout status						
Ever dropped out	95	6	6.3	1	16.7	1.1
Never dropped out	944	97	10.3	21	21.6	2.2
Prior-round response status						
ELS:2002/06 respondent	771	69	8.9	18	26.1	2.3
ELS:2002/06 nonrespondent	268	34	12.7	4	11.8	1.5

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Debriefing Meeting. At the conclusion of the ELS:2002/12 data collection, project staff held a debriefing meeting with interviewers to learn more about their experiences. The interviewers reported positive overall experiences working on the study and provided suggestions for improving full-scale data collection. With regard to training, interviewers suggested additional practice on the CATI-CMS, occupation coder, and refusal conversion. Based on their interactions with sample members and other contacts, interviewers also provided examples of successful techniques for gaining cooperation to incorporate into training for full-scale data collection. Interviewers reported that they received the resources necessary to administer the interview successfully, and identified specific questions that sometimes presented challenges for survey respondents. Interviewers were generally appreciative of the support they received through monitored interviews and quality circle meetings. Project staff prepared a summary of the debriefing meeting for consideration when planning the full-scale survey.

3.3 Results of the Propensity Modeling Experiment

Background of Experiment Approach. To develop new approaches to improve survey outcomes, the ELS:2002/12 field test experimented with methods to address response rate and bias, not only at the weighting adjustment phase, but prior to and during data collection as well. Modeled on responsive design methodologies developed by Groves and Heeringa (2006), the initiative attempts to move beyond relying mainly on the survey response rate as an indicator of data quality. Response rates, at best, provide inconsistent information about bias because they provide little information about who is responding. Although high response rates remain desirable, the ultimate goal of any survey data collection is to provide efficient and unbiased estimates of population parameters. Such an approach is new to ELS:2002 (such methods were not employed in the earlier rounds), and reflects a wider NCES initiative, encompassing for example, the postsecondary longitudinal studies and HSLS:09 as well.

Traditionally in survey data collections, potential bias has been addressed through weighting adjustments after the conclusion of the data collection period. However, using responsive design, the methodology attempts to determine whether the risk of bias can be effectively addressed prior to data collection. Using propensity models to predict response in the second follow-up field test, sample members with a low likelihood of responding (low propensity) were identified prior to the third-follow-up field test data collection and were subsequently targeted for special interventions in an effort to reduce the potential overall bias introduced into final survey estimates.

The most important assumption evaluated in the field test was whether low-propensity cases are fundamentally different from high-propensity cases, particularly in the way they respond to survey questions. If differences in estimates between low- and high-propensity cases do exist and are large enough, survey estimates are likely to be affected. Thus, low-propensity cases could contribute to nonresponse bias.[9] Therefore, bringing more low-propensity cases into the respondent pool could have several additional advantages. First, more low-propensity cases means an expanded and richer imputation donor pool for traditionally underrepresented cases. Second, more low-propensity cases being interviewed could result in more precise weights simply because more data are collected. Finally, targeting resources strategically to cases likely to yield more value in terms of bias is an efficient approach to data collection.

With this in mind, the goal of the approach was to reduce nonresponse bias by employing a different methodology for low likelihood of response cases prior to the start of data collection. The experiment implemented in the field test evaluated the ability to identify low-propensity cases a priori, the effectiveness of an increased cash incentive on low-propensity cases, how survey responses may differ between high- and low-propensity cases, and how low-propensity cases may contribute potential bias in final estimates.

Experimental Process. To assign a case's response propensity for the third follow-up field test, using a logistic regression model we estimated their response propensity in the second follow-up field test by using the sample member's known ELS:2002 second follow-up field test response status as the dependent variable. As independent variables, a range of information known for all respondents and nonrespondents from each prior wave of the longitudinal field test including information from batch tracing activities was examined for significance. The following variables were considered as predictors of a field test sample member's second follow-up field test response outcome:

- base-year response status;
- first follow-up response status;
- whether the respondent ever refused;
- whether the respondent ever scheduled an appointment to complete the interview;

[9] For a more extensive explanation of the relationship between a case's response propensity and nonresponse bias, please see Peytchev et al. (2010).

- whether the respondent was classified as hard to reach;
- the number of calls made to the respondent in the second follow-up;
- high school completion status;
- parental level of education;
- high school type;
- urbanicity;
- dropout status; and
- the sample member's postsecondary aspirations.

Predicted probabilities derived from the logistic regression model were then used to indicate response propensity for the third follow-up field test. The median response propensity was .87 (range = .05 to .96.). To ensure that enough cases were assigned to the treatment and control groups, cases were split into two groups of equal size. Sample members above the median response propensity were classified as high propensity, and those below the median as low propensity. Of the total 1,056 members, 529 cases were classified as high propensity and 527 as low propensity[10]. In addition, the 527 low-propensity cases were randomly split into experimental and control groups. The goal was to examine how well low-propensity cases using prior-wave data can be predicted and how these should be treated to change their response propensity.

Note that the propensity model-based approach utilized in the field test differs from more typical and familiar uses of propensity score matching in other fields. The approach tested in the field test identifies low propensity cases through modeling. Next, treatment and control cases are randomly selected from within the low propensity group, and finally the data are analyzed for experimental effects. In contrast, typical applications of propensity score matching use propensity scores to match participants and non-participants to identify treatment effects. Typically, propensity score matching is used when randomization is not possible, as is the case with many observational studies.

Because low-propensity cases assigned to the experimental group received an increased incentive, of interest is how those cases were distributed according to their prior response status. Table 11 shows the distribution of sample members across propensity groups by selected factors.

As shown in table 11, high-propensity cases were not limited to second follow-up respondents, and low-propensity cases included both respondents and nonrespondents. In fact, a number of nonrespondents were classified as high propensity. This suggests that ELS:2002 prior-round response status, while important, may not be sufficient as a predictor of response outcome in the third follow-up and should not be the sole basis for partitioning cases into propensity categories.

[10] The field test sample size was 1,056 with 17 exclusions (see section 3.2) resulting in 1,039 cases pursued in the third follow-up field test data collection.

Table 11. Distribution of third follow-up propensity cases by prior response status: 2011

Propensity group	By response status					First follow-up response status					Second follow-up response status				
	Respondent		Nonrespondent			Respondent		Nonrespondent			Respondent		Nonrespondent		
	N	Percent	N	Percent	Total	N	Percent	N	Percent	Total	N	Percent	N	Percent	Total
Low-propensity cases	448	85.0	79	15.0	527	411	78.0	116	22.0	527	292	55.5	235	44.6	527
High-propensity cases	508	96.0	21	4.0	529	528	99.8	1	0.2	529	489	92.4	40	7.6	529
Total	956		100		1,056	939		117		1,056	781		275		1,056

NOTE: This table includes ineligibles because propensity assignment and subsequent randomization was done prior to data collection.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Table 12 shows the distribution of the case propensities across some demographic characteristics of interest. From the data in table 11, the overall sample percentages were similar in the high-propensity groups and in the overall sample. There is no obvious over- or underrepresentation in any one subgroup of cases.

Treatment for Low-Propensity Cases. The basic premise of the response propensity approach is to identify low-propensity cases prior to data collection and assign to them a treatment, such as a cash incentive, to encourage response. The treatment for ELS:2002/12 field test low-propensity cases was a higher incentive of $45 at the start of data collection (weeks 1–9) that increased to $55 at week 10. High-propensity and control group cases were offered $25 until week 10 of data collection, when the incentive increased to $35. Table 13 outlines the timing and levels of the different incentives.

Results of the Experiment. The predictive model developed ahead of the field test data collection successfully informed the eventual response outcome for sample members. As shown in table 14, the response rate of the high-propensity group (67 percent) was significantly higher than that of the low-propensity control group (45 percent) ($\chi^2 = 34.9$; $p < .0001$).

Table 12. Number of cases and percentage of completed interviews by data collection phase: 2011

	Cases in field test sample		Cases in high-propensity group	
	Number	Percent	Number	Percent
Total	1,056		529	
Gender				
Male	531	50.3	250	47.3
Female	525	49.7	279	52.7
Race				
White	524	49.6	280	52.9
Black	178	16.9	83	15.7
Hispanic	195	18.5	78	14.7
More than one race	101	9.6	54	10.2
Asian	58	5.5	34	6.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Table 13. ELS:2002 field test treatment schedule

	High response propensity	Low response propensity	
Week	All high cases	Control group	Experimental group
1–9	$25	$25	$45
10+	$35	$35	$55

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Table 14. Response rates by propensity and experimental group

	Percentage respondents	N	Percentage nonrespondents	N	Total
Total eligible					1,039
High propensity	67.4	351	32.6	170	521
Low propensity (control)	45.4	118	54.6	142	260
Experimental group	51.6	133	49.5	125	258
Control group	45.4	118	54.6	142	260

NOTE: Of the 1,056 field test sample cases, 1,039 were eligible for the third follow-up collection. See section 3.2 of the report for information on eligibility.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

In examining the effect of the higher incentive treatment for low-propensity cases, a numerical difference in participation (52 percent for treatment cases and 45 percent for control cases) was observed. However, the difference was not statistically significant ($z = 1.41$, $p = .08$).

One of the analytical goals of the field test was to determine whether high- and low-propensity cases in ELS:2002 answer survey questions differently (i.e., exhibit item-level differences). Following is a list of select survey variables likely of high value to analysts showing significant differences between propensity groups ($p < .05$). Their estimates can be seen in table 15.

- Whether respondent has earned a regular high school diploma
- Whether respondent has attended a postsecondary institution
- Highest level of postsecondary school attended
- Postsecondary attainment (specifically, postsecondary attendance with no credential, and bachelor's degree)
- Educational expectations (specifically, 1- or 2-year degree, master's, and PhD)
- Current marital status (specifically, partnered/living with a significant other)
- Whether respondent is currently registered to vote

Table 15. Estimates of select survey variables by propensity group ($p < .05$): 2011

Group	Low propensity (percent)	High propensity (percent)
Has a regular high school diploma	91.0	95.8
No postsecondary attendance	12.4	6.2
Highest postsecondary institution attended is a 4-year	53.7	64.3
postsecondary attendance but no credential	44.2	29.4
Highest postsecondary credential earned is a bachelor's	26.6	39.9
Highest level of education expected: complete 1- or 2-year program	17.5	10.9
Highest level of education expected: Master's	22.8	30.5
Highest level of education expected: PhD, professional doctorate	8.9	16.1
Current marital status is living with significant other	13.7	7.5
Currently registered to vote	79.8	89.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Other variables showing significant differences between propensity groups ($p < .05$) include the following:

- Currently taking courses at voc/tech/trade school
- Currently working and attending postsecondary school
- Debt/asset ratio is in highest quartile

Other variables with differences between propensity groups approaching significance ($p < .1$) include the following:

- If sold assets and paid debt: break even
- Debt/asset ratio is in middle two quartiles
- Voted in presidential election
- Voted in state or local election

Variables showing no significant differences between propensity groups include the following:

- Currently working at full-time job
- Currently working at part-time job
- Currently taking courses at 2- or 4-year college
- Not currently working and not currently attending postsecondary school
- Currently working but not attending postsecondary school
- Currently attending postsecondary school but not working
- Timing of high school credential
- Taken out any education loans
- Amount owed on education loans
- Number of hours working per week
- Ever unemployed since January 2008

- Total number of months unemployed
- Ever in the military
- Whether respondent has biological child
- Frequency of volunteering

Mean relative bias on the unit level was also examined, as shown in table 16. It was calculated by finding the simple average of a set of relative biases. For the field test analysis, 23 relative biases were calculated from seven variables (sex, race, ever dropout status, high school completion status, base-year school type, base-year urbanicity, and base-year school FIPS state code), and the mean relative bias was calculated by calculating their simple average.

Table 16. Unit-level bias analyses: 2011

Group	Mean relative bias
All	10.4
All with low-propensity cases treated as nonrespondents	17.6
Propensity group	
High propensity	11.1
Low propensity	8.4
Experiment condition	
Low-propensity control	13.7
Low-propensity treatment	10.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up (ELS:2002/12) Field Test.

Bias is defined as the weighted absolute difference between an estimate calculated using respondents only and an estimate calculated using nonrespondents only. The weight for a bias is defined as the nonresponse rate associated with the bias calculation. A relative bias is calculated by dividing the bias by the estimate calculated using both respondents and nonrespondents.

Propensity Modeling Conclusions. The response propensity model successfully predicted response outcome. The inclusion of low-propensity cases showed an apparent reduction in unit-level biases. Including more low-propensity cases in the data may reduce bias and may help improve final estimates because low-propensity cases appear to be different in terms of their survey responses. Though the results were not statistically significant, a higher incentive amount produced a numerically higher response rate for low-propensity cases. It is not known how the experiment would have concluded if the field test data collection had continued 2 additional months to its scheduled end. The data collection ended two months early because the interview yield goal was met. If data collection had continued, more cases would have been traced and located, and more cases would have been successfully reached and interviewed.

There are important lessons learned for the full-scale study. First, from an operational standpoint, strategic case targeting can be implemented well for ELS:2002/12. Interventions were effectively delivered to targeted cases providing confidence that this or a similar process could also be implemented successfully on the much larger full-scale study. However, this field

test experiment revealed only small differences in measured bias; therefore, propensity to respond may not be the best method for strategic case targeting. New approaches for bias reduction should be considered for the full-scale study. New statistics, such as the Mahalanobis distance function, within a responsive design, should be considered for the full-scale study.

Although case propensity may end up not providing sufficient reductions in bias in the ELS:2002/12 data, the process of strategic case targeting for the purpose of improved data quality will be implemented in the study for several important reasons. Prior research has found that nonresponse bias is related to how and with whom new interviews are obtained during nonresponse follow-up (Peytchev et al., 2009). Specifically, data collection actions during nonresponse follow-up can be detrimental to data quality by bringing in sample members who resemble those most likely to respond (Schouten et al., 2009). In the full-scale study, the goal for data collection will be to conduct nonresponse follow-up to maximize the chances of a fully representative respondent pool. Chapter 6 outlines full-scale study recommendations based on the results discussed above.

This page left intentionally blank.

Chapter 4. Field Test Questionnaire Timing and Data Quality

This chapter presents findings from analyses of questionnaire timing, item timing, item frequency distributions, "please specify" text, reliability reinterviews, and the reliability and dimensionality of scales formed from sets of items. In addition to the findings, recommendations for the full-scale study are included. All recommendations may not necessarily be implemented, owing to survey constraints and National Center for Education Statistics (NCES) priorities.

4.1 Questionnaire Timing and Item Analyses

4.1.1 Questionnaire Timing

The Education Longitudinal Study of 2002 (ELS:2002) third follow-up questionnaire repeats some items from prior ELS:2002 rounds, but also contains new items that ask about educational, occupational, and income status. The goal of the field test was to present a questionnaire that would take about 35 minutes on average to complete. As shown in table 17, the average time spent on the questionnaire was 39 minutes, with a median of 36.5 minutes.

In the third follow-up field test, there were 602 participants. Of the 602, 595 completed the survey and 7 partially completed the survey. Approximately 71 percent of the 595 field test respondents who completed the questionnaire finished the questionnaire in 45 minutes or less, and 95 percent of the respondents finished in 66 minutes or less. These results prompted RTI, in conjunction with NCES and with advice that was solicited from the ELS:2002/2012 Technical Review Panel (TRP), to cut items from the field test instrument in hopes of producing a full-scale questionnaire with an average time-to-complete of 35 minutes.

Despite the lengthy time spent on the questionnaire by some respondents, only 7 out of the broader number of 602 field test respondents (1.2 percent) responded to some of the survey but did not reach the end. However, 16 additional respondents (2.7 percent of the 595 respondents) failed to answer more than half of the available items, for a total of approximately 4 percent of field test respondents with significant or completely missing information.

Table 17. Questionnaire timing of ELS:2002 third follow-up field test questionnaire, by response mode and questionnaire section: 2011

Response mode and questionnaire section	Mean minutes	Standard deviation of minutes
Overall (n = 595)		
Total	39.3	15.70
Section A	11.2	55.50
Section B	11.2	5.96
Section C	5.0	2.56
Section D	11.5	5.42
Web (n = 493)		
Total	38.8	16.51
Section A	11.1	6.11
Section B	11.1	6.31
Section C	5.0	2.72
Section D	11.1	5.69
Computer-assisted telephone interviewing (n = 102)		
Total	41.6	10.25
Section A	11.7	4.84
Section B	11.4	3.85
Section C	5.0	1.54
Section D	13.4	3.25

NOTE: Incomplete questionnaires are not included.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up Field Test (ELS:2002/12).

No significant differences in the average completion time were found in the overall questionnaire or in any section when comparing by mode of administration (online or by computer-assisted telephone interview [CATI]) (table 1). Questionnaire timing therefore does not suffer from mode effects and does not require adjustments specific to the mode of administration.

4.1.2 Item Timing

In addition to the overall timing of the instrument by section and mode, the examination of individual item timing data revealed a few areas of concern. Although most items (having two to five response options) averaged 8 to 12 seconds to answer, some took longer. The items that took longer were not clearly a result of lengthy or complex response options or intensive recall requirements. Lengthy items increase the overall survey time and they also uncover potential problems in question and response option wording or design.

After analysis of the item timing data, the following issues were noted for further consideration by the instrumentation team:

- F3JOBDESCRB—Although a relatively simple question ("Which one of the following four statements best describes your job?") with only four response options,

field test respondents averaged 26 seconds answering it. The question may be conceptually difficult because it asks respondents to judge various combinations of who controls "what" and "how" they perform their job. Nevertheless, the TRP was generally supportive of including this item in the full-scale questionnaire as a measure of career-goal realization; and, given that this item was borrowed from the fourth follow-up of NELS:88, F3JOBDESCRB will be retained as is for the full-scale questionnaire to preserve potential trend analyses.

- F3MLTCURBRCH—This item asks about the current military branch in which a respondent serves, if he or she indicated service in multiple branches. With only four response options, respondents took an average of 80 seconds to answer. However, this question only applied to two respondents, so the lengthy response is likely attributable to a single case (although more respondents would be expected to answer this item in the full-scale sample). After discussing this item at the post–field test TRP meeting, this item was deemed to be of relatively lesser importance and will not be included in the full-scale instrument.
- F3RFUND—Asking about retirement savings plans, this item had five response options and averaged 24 seconds. The question wording may contribute to the lengthy time, given the long question stem that distinguishes between employer-based and individual retirement plans and lists seven specific types of plans. Additionally, the response options are not aligned with the question's emphasis on the distinction between employer and individual plans. Instead, the response options distinguish personal plans from spouse or partner plans and ask about different combinations of personal and spouse/partner planning—including introducing the term "joint plan" in the second response option before it has been defined in the third response option. Furthermore, a response option appears to be missing (spouse/partner has own plan as well as joint plan with respondent). To address the problems associated with the field test response options, this question will be recast in the full-scale instrument (when administered to respondents who are currently married or living with a significant other) as a series of three more-appropriately ordered yes/no questions: "Do you and your [spouse/partner] have any of these plans jointly?"; "Do you have any of these plans on your own separate from your [spouse/partner]?"; and "Does your [spouse/partner] have any of these plans separate from you?" Full-scale survey respondents who are not currently married or living with a significant other will simply be asked, "Do you have any savings in these types of plans?"
- F3DONATE—Asking about the frequency of voluntary contributions, this item averaged 20 seconds despite five response options presented as a 1–5 scale ("never" to "always"). The question is unclear about the inclusion of religious organizations and donations of goods. Also, the never/always scale is not a concrete time span (i.e., annually, monthly, or weekly) and has no connection to financial amounts, which is a common metric of financial contributions. This item (borrowed from the UniLOA [University Learning Outcomes Assessment]) was discussed at the post–field test TRP meeting, and was ultimately retained for the full-scale instrument because of (1) the promising distribution of responses it produced during the field test; (2) the strong psychometric properties of UniLOA items; (3) its relative brevity; and (4) the lack of another suitable well-tested item to use in its place.

4.1.3 Item Response Distribution

We examined the distribution of responses for each item to identify response options that were chosen by respondents very infrequently or nearly universally. These items provide limited discriminatory information and are candidates for revision or removal. We also examined patterns of missing data to identify items where further encouragement of respondents or alternative placement in the questionnaire might be helpful.

The *minimum distributional criterion* for any item is normally that it exhibits nonzero variability.[11] Only one item (F3EVERADOPT) did not possess this characteristic. This item asks whether the respondent had ever adopted a child, and no respondents indicated "yes." Because nonzero variability is likely for this item in the main study sample of ELS:2002/12 respondents and subsequent questions on adoption would also need to be removed, this item was retained in the full-scale instrument.

Even though all items (except F3EVERADOPT) exhibited variability, some response options within items were chosen at low frequencies (including not at all), which calls into question their analytical utility. To assess low-frequency responses, we examined the field test instrument for response options chosen by 2 percent or fewer respondents (about 12 cases out of the 595 field test respondents).

The frequency of some responses is expected to be low, such as items that asked about the month and year of a particular event, questions about monetary amounts (e.g., earnings), and items asking for a count of events or counts of family members. Because individual values would be rare, these types of items were excluded from the analysis. Table 18 contains the items with the low-frequency response options along with the percentage of field test respondents answering and the specific number of field test cases.

Several of the first items listed in table 2, although they have some responses that are low frequency, are important to keep in the survey for completeness. This includes items or response options about current military service (F3ACTJUNE8), type of high school credential earned (certificate of attendance under F3HSCRED), and type of postsecondary credential earned at the last postsecondary institution attended (PhD or equivalent under F3PSCREDTYPE). Likewise, F3OTHCREDTYPE_1, although it has missing responses, is part of a series of items asking about credentials other than that earned at the last postsecondary institution, and should be kept as is for that reason.

[11] There are reasonable exceptions, particularly in a longitudinal study. If one is looking for the change of some phenomenon over time, it is appropriate to repeat a prior question even when most respondents will not be expected to report it. For all practical purposes within ELS:2002/12, however, nonzero variability is a legitimate minimum standard.

Table 18. Low-frequency responses to ELS:2002 third follow-up field test questionnaire items: 2011

Variable	Description	Response option with low frequency	Number	Percent
F3ACTJUNE8	Serving in armed forces in June 2011	Yes	2	1.8
F3HSCRED	Type of high school credential received	Certificate of attendance	1	0.9
F3PSCREDTYPE	Type of credential earned at last postsecondary institution	PhD or equivalent	1	0.3
F3OTHCREDTYPE_1	Other type of credential earned	PhD or equivalent	0	0
		Professional doctorate	1	0.7
F3CREDSAT1	Satisfaction with faculty	Very dissatisfied	10	1.8
F3WHYNOCRED13	Reason why did not complete postsecondary credential—School or program closed	Yes	2	1.8
F3JOBDETL1	Confident can perform job well	Disagree	1	0.2
		Strongly disagree	2	0.4
F3JOBDETL2	Confident can solve big problems at work	Disagree	2	0.4
		Strongly disagree	2	0.4
F3JOBDETL3	Confident can reach larger goals at work	Disagree	5	1.0
		Strongly disagree	2	0.4
F3JOBDETL4	Certain can do work despite pressure	Disagree	3	0.6
		Strongly disagree	2	0.4
F3JOBDETL5	Certain can juggle work/non-work	Disagree	5	1.0
		Strongly disagree	4	0.8
F3JOBREMAIN1	Job gets respect from friends/family	Strongly disagree	8	1.7
F3OCC30NOT	Does not plan to be working at age 30	Yes	11	1.8
F3MARSTAT	Marital status	Separated	10	1.6
F3MARPARTNER	Ever been legally married (for those in marriage-like relationship)	Yes	1	1.6
F3EVERADOPT	Ever adopted a child	Yes	0	0
F3PARHOME	Lives in parent's home	0 (coding error)	3	1.7
F3INCOMSIN6	Received income from veteran's benefits	Yes	7	1.9
F3INCOMMAR_R_4_1	Received income from Social Security	Yes	2	0.8
F3INCOMMAR_R_4_2	Spouse received income from Social Security	Yes	3	1.2
F3INCOMMAR_R_5_1	Received income from child support	Yes	2	0.8
F3INCOMMAR_R_5_2	Spouse received income from child support	Yes	3	1.2
F3INCOMMAR_R_6_1	Received income from veteran's benefits	Yes	1	0.4
F3INCOMMAR_R_6_2	Spouse received income from veteran's benefits	Yes	3	1.2
F3INCOMMAR_R_11_1	Received any other taxable income	Yes	2	0.8
F3INCOMMAR_R_11_2	Spouse received any other taxable income	Yes	3	1.2
F3LIFEEVENT1	Parents or guardians separated	Has happened more than once	2	0.3
F3LIFEEVENT4	One of parents has died	Has happened more than once	3	0.5

See notes at end of table.

Table 18. Low-frequency responses to ELS:2002 third follow-up field test questionnaire items: 2011—continued

Variable	Description	Response option with low frequency	Number	Percent
F3LIFEEVENT8	Victim of violent crime	Has happened more than once	11	1.8
F3VALUES1	Importance of being successful in work	Not important	10	1.6
F3VALUES5	Importance of finding steady work	Not important	11	1.8
F3VALUES12	Importance of leisure time	Not important	7	1.1

NOTE: The following types of follow-up items are not included in the tabular review of low-frequency items because of a large number of response options: items asking for month and year of some event; state in which event occurred; specific institution names; highest level of education expected to complete; specific dollar amounts or wages; number of certain events such as jobs or children; number of hours or weeks for jobs or training.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up Field Test (ELS:2002/12).

A series of newly constructed items designed to capture information on six social-cognitive career theory constructs were included in the field test instrument; a subset of those items (F3JOBDETL1-5 and F3JOBREMAIN1) have low-frequency responses on the disagreement pole of their Likert response options. Nearly all of the responses for the JOBDETL items are in agreement with the statements, with even the “neither agree nor disagree” middle option being chosen typically by only about 5 percent of field test respondents. It may be that the framing of these questions as “confidence” or “certainty” promotes overestimation of respondents’ abilities or a bias toward providing desirable answers to the survey. The F3JOBDETL series of items will not be included in the full-scale instrument, due in large part to this skewed field test response distribution. The JOBREMAIN series, asking about the personal and social benefits of the respondent’s job, is tilted to the positive axis of the response options, but not to nearly as great a degree as the JOBDETL series; for example, the neutral responses are quite frequent. However, factor analyses of field test data (see also section 4.4) indicate that the F3JOBREMAIN series tended to load on the same underlying factor as two other item sets (F3JOBST1-3 and F3JOBST7-9), and thus the F3JOBREMAIN series will not be included in the full-scale instrument either.

F2OCC30NOT (whether respondent plans to work at age 30) and F3MARSTAT (marital status) have low-frequency responses, but the former is a critical gate for a follow-up question about occupational expectations for age 30, and the latter must provide comprehensive options. Likewise, only one field test respondent answered ‘yes’ to F3MARPARTNER, a question which asks about ever having been legally married for those answering to the previous marriage status question, “partnered with significant other.” This item will not be included in the full-scale instrument, however, because current marital status will be obtained by a slightly different line of questioning, (i.e., “partnered with significant other” will be dropped as a response option for F3MARSTAT, but an explicit “Are you currently living with a significant other?” follow-up question will be administered to all full-scale respondents, except those who are currently married).

Several items seem to add little to the questionnaire. A question about why a respondent did not complete a postsecondary degree (F2WHYNOCRED13—because of school or program closure or lost accreditation) had very few positive responses. This item will not be included in the full-scale questionnaire because it does not add sufficient information to the already-long list of other reasons (F3WHYNOCRED1-12) that the questionnaire asks. F3PARHOME, asking about living in parent's home, had one low-frequency response that is a coding error.

Nine of the low-frequency items ask about income received from various sources by the respondent or the respondent's spouse (e.g., F3INCOMSIN6 and the F3INCOMMAR_R series—the F3INCOMMAR_R series is administered to respondents who are married or partnered, while the F3INCOMSIN series is administered to all other respondents). Income received from Social Security and from veterans benefits, both of which are expected to be rare at this age range, are uncommon in the field test responses, as is income from child support and unspecified other income.

The F3INCOMMAR_R series has an additional problem; that is, it not only separately asks if the respondent or the respondent's spouse/partner received income from each source, but also asks whether "neither" the respondent nor spouse/partner received income from each source (not all of the F3INCOMMAR_R items are shown in table 2 because frequencies were not low for some of the items). This involves double-questioning on the same issue, because a respondent who failed to affirmatively check, for example, either the "respondent" or "spouse/partner" box on the form is implying that neither received such income. As a result of these difficulties, the parallel sets of field test items F3INCOMSIN and F3INCOMMAR will not be included in the full-scale instrument.

The final two sets of questions with low-frequency responses involve stressful life events and the importance of specific goals or values. Both series of items are part of longer series in which broader distributions of responses occur. In addition, for the life event items, distinguishing "happened once" from "happened more than once" serves to route respondents to a subsequent set of questions about, respectively, when the single event happened or when the first and then the most recent event happened. Given this integration with other items, no changes were made to the field test questionnaire when producing the full-scale questionnaire. The values questions are likewise embedded in a series involving broader response distributions, and are also part of a series that has been asked in prior rounds of ELS:2002, as well as in the National Education Longitudinal Study of 1988. Although 3 of the 14 items in this series will be dropped from the full-scale instrument in hopes of shortening the overall average interview length, the remaining 11 "value" items will be preserved to maintain longitudinal and cross-cohort comparability.

As a final note for this section, even though it does not result in low response rates, it should be noted that the field test response categories for a question about highest degree ever expected (F3EDEXP) do not match those for a later question about the degree needed for the job

the respondent expects to hold at age 30 (F3OCC30ED). This not only precludes direct comparison between those responses—an obvious analytical use of the data—but may even confuse respondents to the questionnaire. Thus, these response categories will be aligned in their corresponding full-scale versions.

4.2 Items With "Please Specify" Text

Two types of items on the questionnaire allowed respondents to write a text response as an answer. The first type was a methodological probe that began with a yes/no question about whether the respondent had trouble understanding or answering the prior question or set of questions. If yes was chosen, the respondent was prompted to detail the difficulty he or she had. The second type was within a series of items that provided "other" as the last item in order to be exhaustive. If "other" was chosen, the respondent was asked to specify to what he or she was referring. Examining these text responses provides a way of confirming that existing options adequately cover the likely responses and that question stems and response options are understood. Table 19 presents the percent and number of cases that had difficulty or used an "other" option.

Table 19. Items with "please specify" text: Methodological probes and items with an "other" option: ELS:2002 third follow-up field test: 2011

		Had difficulty or "other" response	
Item	Description	Number	Percent
	Methodological probes (had difficulty answering item)		
F3LOANPAID1-3	Student loan debt paid by respondent/paid by family/forgiven by a loan forgiveness program (three response options)	5	9.8
F3LOANPAIDALT1-3	Student loan debt paid by respondent/paid by family/forgiven by a loan forgiveness program (two response options)	16	5.9
F3LICENSE	Has state or industry license or professional certification	20	3.4
F3JTRAIN	Received formal job training from employer or union in last 12 months	19	3.4
F3BIOCHPART	Was married or partnered with child's parent at time of child's birth	8	4.8
F3PARHOME	Respondent lives with parents	3	1.7
	"Other" response options		
F3LOANAFFECT5	Student loan debt influenced employment decisions or plans in some other ways	58	19.0
F3VOLORG9	Unpaid volunteer or community service work—other community or service work	41	18.0

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up Field Test (ELS:2002/12).

4.2.1 Methodological Probes

Six items involved methodological probes. The percentage of respondents who indicated they had problems answering a question or question series ranged from 2 to 10 percent, and was below 5 percent for four of the six items. The first two item series in table 19 had the largest percentage of respondents indicating a problem, and they were alternate versions of the same

question. For both, the question stem was "Has any of your student loan debt been…", and individual items were "paid off by you," "paid off by your family," and "forgiven by a loan forgiveness program?" For the first item series (F3LOANPAID), response options were "none of the debt," "some of the debt," and "all of the debt." This item was presented to respondents who owed nothing on their loans. For the second item series (F3LOANPAIDALT), response options were "none of the debt" and "some of the debt" only. This item was presented to respondents who still owed money on their loans. A greater percentage of respondents indicated that they had difficulty with the first item (10 percent) than the second item (6 percent). The "please specify" text indicated that a number of respondents with loan amounts (who therefore received F3LOANPAIDALT) said they were still enrolled or were within the grace period before repayment was required, and therefore did not have to pay anything. However, the questionnaire routes respondents to F3LOANPAIDALT even if they indicated on a prior question (F3LOANPAY) that they pay nothing each month. Because of the issues described here, and because several questions on debt are included later in the instrument, neither version of these items (F3LOANPAID or F3LOANPAIDALT) will be retained in the full-scale instrument.

The other four methodological probes covered job and family issues. For the question about professional certification and state and industry licensure (F3LICENSE), 3 percent of field test respondents indicated a problem answering. The "please specify" text indicated that some respondents had difficulty understanding what was meant by a "professional certification" or a "state or industry license" (for example, one asked if a teaching license was included). We recommend adding examples to this question, such as teaching license or computer certification.

For the job training question (F3JTRAIN), 3 percent of respondents indicated a problem answering. The "please specify" responses did not indicate any consistent issue, but this item was used in the reliability reinterview discussed in section 4.3 below (the discussion there critiques the lengthy, confusing wording of this item). For the question about marital or partner status when the respondent's child was born (F3BIOCHPART), 5 percent indicated a problem answering. There were only three responses supplied to the "please specify" request, and they did not reveal a consistent problem. One respondent indicated some confusion about the meaning of "partnered with," because that person was not married but was living with the child's other parent. For the full-scale instrument, "partnered with" will be removed from the question wording because existing research indicates not being married (as opposed to not being married or partnered) is the primary distinction in determining whether having a child might interfere with other education/employment-related activities. For the question about a parent living with the respondent (F3PARHOME), fewer than 2 percent of field test respondents indicated a problem, and only one vague response to "please specify" was provided; therefore, no changes to this item will be made for the full-scale instrument.

4.2.2 "Please Specify" Responses at End of Series

Two items involved providing text for "other" responses at the end of a series of items. F3LOANAFFECT5 was an "other" response for ways in which student loan debt affected employment decisions, while F3VOLORG9 was an "other" response for types of volunteer or community service work the respondent had performed. The percentage of respondents choosing "other" was substantial in both cases: 19 and 18 percent, respectively. The "please specify" text for loan effects indicated that a number of respondents could have chosen existing response options. For example, six respondents indicated that they had to work multiple jobs, yet presumably this would have been covered by an existing option, "You had to work more than one job at the same time." However, this option does not allow for multiple types of jobs or serialized work in different areas.

The responses to the "other" option on the volunteering question indicated sets of respondents who worked for animal welfare organizations (six respondents), fire departments (five respondents), and other medical groups or causes (nine respondents). These responses suggest that the existing option for "hospital or nursing home" should be broadened to something like "health care or medical cause." In addition, either additions should be made for "animal welfare" and "public safety (including volunteer fire departments)" or examples should be added to existing options (e.g., to "school or community organization" or "neighborhood or social action association"). However, another option—and the option that was implemented for the full-scale questionnaire—was to replace this list of volunteer organizations with one from cognitive testing, that covers a wider array of, in particular, health organizations.

4.3 Reliability Reinterview: Design and Results

A subsample of respondents was selected at random to complete a reinterview designed to assess the consistency of selected questions. The following section summarizes results from an analysis of these reinterviews.

Reinterviews were conducted in CATI or in web self-administration at least 2 weeks following the completion of the first interview. By the end of data collection, 60 respondents had completed a reinterview. The reinterview consisted of questions which applied to all respondents. Items were not selected for reinterview if, owing to skip patterns, not enough respondents would be administered the item to yield sufficient data for analysis. Thirty-three items were selected for the reinterview—25 categorical items (mostly yes or no items) and 8 continuous items.

Results of the analysis of the reinterview items are displayed in two tables (4 and 5). In both tables, values are based on cases where a response was provided in both interviews. Values of 85 percent agreement or above are considered as exhibiting high reliability, between 55 percent and 85 percent as moderate reliability, and below 55 percent as low reliability.

Categorical Items. Table 20 presents the categorical items with response options of yes/no or other two-category responses. It shows percentage agreement; Cramer's V, a measure of the strength of association, which ranges from 0 to 1, with 1 indicating perfect association; and the statistical significance of the association between the original interview and reinterview, where cell sizes were sufficiently large to produce valid chi-squared statistics.

Table 20. Interview-reinterview agreement for categorical items on the ELS:2002 third follow-up field test reliability reinterview: 2011

Item	Percentage agreement	Cramer's V	Statistical significance[1]
F3ACTCURR1-8:			
Currently working for pay for 35 hours per week or more	90.9	0.81	*p* <.0001
Currently working for pay for less than 35 hours per week	87.3	0.65	‡
Currently serving in another work experience (apprenticeship, training program, or internship)	94.6	0.70	‡
Currently taking courses at a 2-year or 4-year college	92.7	0.84	*p* <.0001
Currently taking courses at a vocational, technical, or trade school	94.6	0.70	‡
Currently keeping house full-time	90.9	0.63	‡
Currently caring for dependent children or adults	98.0	0.95	‡
Currently serving in the armed forces or military	98.2	0.70	‡
Work and school activities in last week of June 2011 same as now (F3ACTSAME)	73.2	0.30	‡
Ever attended college, university, vocational, technical, or trade school since 2005 or since high school (F3EVERATT)	94.1	0.88	*p* <.0001
Lives by him- or herself or with others (F3HHOTHERS)	87.8	0.66	‡
Received formal job training from employer or union in last 12 months (F3JTRAIN)	72.7	0.45	*p* = 0.0009
F3LIFEEVENT1-8:			
Parents or guardians got divorced or separated since January 2005	96.3	0.80	‡
One of parents or guardians lost his or her job since January 2005	85.2	0.85	‡
Respondent lost job since January 2005	83.6	0.61	‡
One of parents or guardians died since January 2005	96.3	0.71	‡
Close relative or friend died since January 2005	75.9	0.59	‡
Respondent became seriously ill or disabled since January 2005	92.7	0.63	‡
Family member became seriously ill or disabled since January 2005	66.7	0.50	‡
Respondent was victim of violent crime since January 2005	98.2	1.00	‡
Current marital status (F3MARSTAT)	91.1	0.86	‡
Performed unpaid volunteer or community service work in past 2 years (F3VOLUNTEER)	87.3	0.74	*p* <.0001
Currently registered to vote (F3VOTE)	92.9	0.74	‡
Voted in a local, state, or national election in past 2 years (F3VOTELOCAL)	85.5	0.70	*p* <.0001

See notes at end of table.

Table 20. Interview-reinterview agreement for categorical items on the ELS:2002 third follow-up field test reliability reinterview: 2011—continued

Item	Percentage agreement	Cramer's V	Statistical significance[1]
Voted in 2008 presidential election (F3VOTEPRES)	96.4	0.92	*p* <.0001

[1] Test of the chi-squared statistic.
‡ Reporting standards not met. Chi square for this item could be invalid; some cells have expected counts of fewer than five.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up Field Test (ELS:2002/12).

For the categorical variables, the percentage agreement ranged from 67 percent to 98 percent, with 20 of the 25 items having matched responses in at least 85 percent of the cases. Five items had percentage agreement between 55 percent and 84 percent, and none had a percentage agreement below 55.

Of the five items with moderate reliability, the first is a question about whether work and school activities at the time of the interview are the same as they were in the last week of June 2011, which is designed (along with follow-up questions) to anchor all of the respondent's activities at the same time point for comparability across cases during analysis. This item had a respectable percent agreement of 72 percent, but a low Cramer's V of 0.30 (the lowest of all reinterview items). Although a simple yes/no question, the anchoring of this question at a specific week, even a recent week, may make recall difficult for some respondents. Alternatively, the number of different items involved in "work or school activities" asked about in this question makes it quite possible that one or more of the activities that the respondent is comparing his or her past to could have changed. For this reason, and the importance of anchoring activities to a specific time point, no changes to this item will be made for the full-scale instrument.

The next moderate-reliability item asks "In the last 12 months, have you participated in a formal training program offered by an employer or a union that helped you learn or improve the skills needed to do your job?" Although the percent agreement is also respectable at 73 percent, and there is a statistically significant association between the original and reinterview responses, Cramer's V is relatively low at 0.45. As another simple factual question, we might expect percent agreement to be higher. However, the 12-month reference window used by the reinterview is by definition slightly different from the 12-month reference window used in the original interview and may account for some of the disagreement. We therefore implemented no changes.

The remaining three items with moderate reliability are part of a series which asks about certain events occurring since January 2005 (about a 5.5-year time span). The items in this series that have high reliability ask about events that the respondent experienced (such as becoming seriously ill) or highly significant events involving parents or guardians (such as death or divorce). In contrast, two of the three moderate reliability items ask about less-significant events

that others experience: close relative or friend died (percent agreement = 76, Cramer's V = 0.59) and family member became seriously ill or disabled (percent agreement = 67, Cramer's V = 0.50). The ambiguity of "close relative or friend" and, particularly, of "family member," may be problematic. The third item in this series with moderate reliability asks whether respondent lost a job (percent agreement = 84, Cramer's V = 0.61), and does not have a clear explanation for its somewhat lower percent agreement, given that it involves both the respondent and is likely a very memorable event. Thus, given that the percent agreement is not particularly low, no changes to this item will be made for the full-scale instrument.

Continuous Items. Because percent agreement is not as meaningful when the number of possible responses is large, results are presented separately for continuous variables (all addressing household membership) in table 21. Table 21 reports percentage agreement; a measure of association (Pearson's correlation coefficient *r*, which ranges from -1 to 1); and the statistical significance of an F statistic from a bivariate regression between the original and reinterviewed responses.

Table 21. Interview-reinterview agreement for continuous items on the ELS:2002 third follow-up field test reliability reinterview: 2011

Item	Percent agreement	Pearson's correlation coefficient (*r*)	Statistical significance (F test)
Number of spouses living with respondent	100.0	1.00	†
Number of marriage-like partners living with respondent	100.0	1.00	†
Number of mothers or female guardians living with respondent	84.9	0.69	F < 0.0001
Number of fathers or male guardians living with respondent	89.7	0.79	F < 0.0001
Number of friends or roommates living with respondent	80.8	0.87	F < 0.0001
Number of brothers or sisters living with respondent	92.3	0.89	F < 0.0001
Number of children living with respondent	96.3	0.94	F < 0.0001
Number of others living with respondent	100.0	1.00	†

† Not applicable
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up Field Test (ELS:2002/12).

Table 21 shows the eight continuous items that were asked in reinterviews. Percentage agreement ranges from 81 percent to 100 percent, with six of the eight items showing agreement above 85 percent. The two moderate-agreement items are attached to questions about the number of mothers or female guardians living with the respondent (just under 85 percent agreement, with Pearson's *r* at 0.69) and the number of friends or roommates living with the respondent (81 percent agreement, Pearson's *r* 0.87). The F tests of bivariate regressions between the original and reinterview responses for both items were statistically significant. Given these relatively good results, and no obvious problems in question construction, no changes to either item will be made for the full-scale instrument.

4.4 Reliability Analysis for Scales

The field test questionnaire included several sets of items designed to capture information on a single construct such as occupational self-efficacy or job satisfaction. By soliciting information about a single construct from multiple items, the construct can be measured more accurately and reliably—as part of a single scale—than a single-item question. For the ELS:2002/12 field test, the items that were intended to form scales were new to the ELS:2002 study, with the scales focusing primarily on occupational issues. These scales were created specifically for ELS:2002 third follow-up by Professor Robert Lent of the University of Maryland, and are based on social cognitive career theory.[12] This section examines the dimensionality and reliability of these scales.

Seven scales were computed from the questionnaire: six scales which cover social cognitive career constructs, and one—subsequently dropped—which covered satisfaction with undergraduate education. All of the scales are constructed as the standardized sum of the respondent's answers to a series of related questions. For each scale, the following measures were computed to assess which scales are recommended for inclusion in the main study: each item's correlation (Pearson's r) with the overall scale, each item's correlation with a scale computed without that item, and the overall reliability (i.e., internal consistency) of the scale (alpha or α, also known as Cronbach's coefficient alpha [Cronbach 1951]). The reliability coefficient α is the square of the correlation between the scale and the underlying dimension or factor; α thus represents the expected correlation of the scale with a scale formed from the same number of alternative items. This section summarizes findings concerning scale reliability; more details about reliability for each of these scales are presented in appendix E.

Factor analysis indicated that each set of items was unidimensional; that is, each set of items was linked to a single factor that could be usefully summarized as a scale score (the large majority of variance within each set of items was accounted for by a single factor for those items, and eigenvalues were large for the first factor in each case and dropped sharply afterward).

Table 22 lists the scales and shows their standardized α. It was the objective of ELS:2002/12 to have moderate to high reliability for all scales, and an α of 0.80 or above was taken as the standard for high reliability, with reliabilities between 0.65 and 0.79 denoting moderate reliability (Nunally and Bernstein 1994).[13] The reliabilities for the seven scales ranged from a low of 0.799 to a high of 0.930. Thus, on the basis of the reliability criteria, all scales were highly reliable, and all are recommended for inclusion in the main study. In addition, based

[12] Social cognitive career theory (for an influential statement of this approach, see Lent et al. [1994]) builds on the work of Bandura and others in social learning theory. Self-efficacy beliefs are seen as interacting with outcome expectations and goals. The theory is frequently applied in the context of career choice and development, including the explanation of subgroup differences.

[13] Nunally and Bernstein (1994) suggest that an internal consistency coefficient of 0.80 is sufficient in research contexts where the purpose of scaling is group-level comparison and research. Where clinical decisions or individual judgments are the main focus, reliabilities of 0.90 and above may be required.

on the results provided in the appendix, no items were recommended for deletion from any scale, on the basis of this initial analysis.

Table 22. Reliability analysis of questionnaire scales: Standardized alpha

Scale name	Reliability (alpha)
Occupational self-efficacy (F3JOBDETL1-5)	0.903
Occupational interest (F3JOBST1-3)	0.899
Support at work (F3JOBST4-6)	0.860
Job benefits (F3JOBREMAIN1-4)	0.799
Job satisfaction (F3JOBST7-9)	0.930
Job commitment (F3JOBST10-12)	0.921
College satisfaction (F3CREDSAT1-4)	0.835

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Education Longitudinal Study of 2002 Third Follow-up Field Test (ELS:2002/12).

Because, however, the six social cognitive career theory scales (i.e., all but the college satisfaction scale) were newly constructed specifically for ELS:2002/12, additional analyses were conducted to ensure that they would prove efficacious. For further evidence that items and constructs were functioning well psychometrically, exploratory and confirmatory factor analyses were undertaken.

Theoretically, these six item sets should have corresponded to six correlated but relatively distinct constructs: occupational self-efficacy, interest, supports, benefits, satisfaction, and commitment. As already noted, all item sets produced acceptable internal consistency reliability estimates. However, when all items were pooled, three of these item sets (interest, benefits [outcome expectations], and satisfaction) tended to load on the same underlying factor. That is, rather than distinguishing between them, participants tended to respond to them as if they were all reflecting a single, larger construct of positive perceptions of one's job. The three other item sets (support, satisfaction, commitment [persistence intention]) tended to load on correlated but reasonably distinct factors.

The distribution of self-efficacy was highly skewed and kurtotic, with most participants feeling highly confident about their ability to perform their job duties and few feeling insecure. This non-normal distribution likely attenuated the correlations between self-efficacy and the other constructs (the correlations were much lower than those typically found in the literature). Support scores were also quite skewed and kurtotic, although support did still yield plausible correlations with most other constructs.

The three social cognitive career theory scales that seem most justifiable to include are support, satisfaction, and commitment. All three lend themselves to trait-like measurement (i.e., they do not need to be linked to specific types of jobs to be assessed adequately). The items which comprise the remaining SCCT scales (self-efficacy, interest, and benefits [outcome expectations]) will not be included in the full-scale instrument.

This page left intentionally blank.

Chapter 5. Survey Control Systems and Data Processing

5.1 System Design, Development, and Testing

Systems and processes used in the Education Longitudinal Study of 2002 (ELS:2002) third follow-up field test were designed and developed to test and identify areas of improvement in preparation for the full-scale study. The smaller scope of the field test provides opportunities to test systems and processes which often can provide greater efficiencies in the full-scale study. The following are the major systems that were used for the field test:

- Integrated Management System (IMS) (a comprehensive tool used to exchange files between RTI and the National Center for Education Statistics (NCES), and to provide access to a centralized repository of data collection reports);
- Survey Control System (SCS) (the central repository of the status of each activity for each case in the study);
- Hatteras Survey Editor (a web-based application used to develop the ELS:2002 third follow-up instrument);
- Web Survey (a web-based application used to administer the ELS:2002 third follow-up instrument);
- Computer-assisted telephone interview Case Management System (CATI-CMS) (a system that assigns cases to be called and provides telephone interviewers with the appropriate screens and scripts to be used during the contacting and locating phase of CATI); and
- ELS:2002 sample member website (public website hosted at NCES and used to disseminate information, collect sample data, and administer the ELS:2002 survey).

Each system is described in detail in this chapter.

Systems were developed using the full system development life cycle process. Each system makes necessary safeguards to handle personally identifying information (PII). Systems such as IMS, Hatteras, and CATI-CMS are standard RTI systems used successfully on other NCES studies (e.g., High School Longitudinal Study of 2009, Baccalaureate and Beyond), and were developed using the latest software tools such as Microsoft.NET and Microsoft SQL Server database management system.

Processing of PII by all systems was developed in accordance with the Federal Information Processing Standards (FIPS) moderate security standard. Movement of the data

containing PII between the locations was handled in accordance with security requirements. In compliance with the FIPS 140.2 standards, data were encrypted when moved between locations, and decrypted once successfully reaching the destination. In addition to security, these automated systems were developed to handle the need of moving data and files efficiently.

5.1.1 Integrated Management System

The IMS is a web-based system that gives project staff and NCES ready access to a repository of reports and other project information and deliverables. The IMS website provides online, instant access to project management tools, such as the current project schedule, monthly progress reports, daily data collection reports and status reports, and project plans and specifications.

5.1.2 Survey Control System

The SCS refers to the control system database and its integrated set of applications used to control and monitor all activities related to data collection, including tracing and locating. Through the control system applications, project staff were able to perform such activities as e-mailing to groups of sample members, preparing lead letters and follow-up mailings, executing batch tracing, reviewing locating information, tracking case statuses, and viewing comments from telephone interviewers. The control system served as a repository of data collection systems needing sample-member-specific data access to a single database. Batch processes pulled information across data collection systems and updated the control system database on a nightly basis.

5.1.3 Hatteras Survey Engine and Survey Editor

Hatteras is the web-based system in which project staff developed, reviewed, tested, modified, and communicated changes to specifications and code for the ELS:2002/12 field test instrument. Hatteras provides tools to conduct the same survey in a multimode (web or CATI) survey instrument. Hatteras provided specification, programming, and testing interfaces for the ELS:2002/12 third follow-up field test instrument.

Survey Editor is an interface for editing the instrument specifications such as question wording, routing of each survey screen, instructions, help text, and item documentation. All information relating to the instrument was stored in a SQL Server database and was accessed through Survey Editor. Once the web survey had been programmed, testers could enter comments into Hatteras, which included a comprehensive comment tracking system to ensure resolution. Hatteras also facilitated importing and exporting item documentation and other information associated with instrument development.

Hatteras allows nonprogramming staff to do much of the specification work for instrument questions and items by automatically translating specifications into web page scripts. For questions involving complex routing, varying question and response content, or unusual page

layout or behavior, programmers entered custom programming code (HTML, Javascript, and C#.NET script) into the Hatteras custom code interface. This code was stored in the SQL Server database along with the instrument specifications for compilation by the survey execution engine.

5.1.4 Web Survey

The Hatteras system's survey execution engine allowed immediate testing of specification and code content as it was entered and updated, displaying web content as respondents would see it. The execution engine also automatically handled such web instrument functions as backing up and moving forward, and recording instrument timing data.

For web and telephone data collection, the Hatteras system was installed on NCES's web server farm and SQL Server database. Web respondents accessed the survey directly by web browser after logging in with a user ID and password. RTI's telephone interviewers accessed the same NCES web survey site by means of a web browser process launched from CATI-CMS. All connections to the NCES web interview were secured with Secure Sockets Layer (SSL) encryption. Automated processes transferred data between the NCES database and the RTI database via a secure, encrypted connection. RTI's database was housed in an Enhanced Security Network (ESN), a separate storage network which is certified to meet the standards required for protection of data classified as FIPS-Moderate as defined by the National Institute of Standards and Technology (NIST)—NIST produces guidelines that relate to levels of information security.

5.1.5 Computer-Assisted Telephone Interview Case Management System

The CATI-CMS is a comprehensive system that manages all aspects of telephone-based data collection. The CATI-CMS connects the various components of the CATI system, including the questionnaire, utility screens, databases, call scheduler, report modules, links to outside systems, and other system components.

The call scheduler, a major tool in CATI-CMS, delivers cases to interviewers in a predefined priority order. In addition to delivering cases with appointments to interviewers at the appropriate time, the call scheduler also calculates the priority scores (the order in which cases need to be called based on preprogrammed rules), sorts cases in non-appointment queues, and computes time zone adjustments to ensure that cases are not delivered outside the specified calling hours. The call scheduler also permits callbacks to be set, and assigns status codes to the case. In addition, each case contains one or more roster lines that detail specific contact information for a case (e.g., home phone number, work phone number). The call scheduler uses a call algorithm based on the previous call results to determine which roster line should be called next.

5.1.6 ELS:2002 Sample-Member Website

The ELS:2002 public website was hosted at NCES servers, which acted as the main source for information to sample members about ELS:2002. The website provided an option for sample members to update their locating information and another option to complete the third follow-up survey.

5.2 Data Processing and File Preparation

Questionnaire data were collected via two modes: self-administered questionnaire and CATI. The interface to these modes of data collection is the same web-based instrument, storing all response data in the same SQL Server database.

All respondent records in the final dataset were verified with the SCS to spot inconsistencies. For example, if data were collected for a respondent who was later set to an ineligible status, the SCS would serve as a safeguard to ensure that the ineligible case was excluded from the final data file. Furthermore, the data files serve as a check against the SCS to ensure that all respondent information is included in production reports. Frequency reviews are conducted on edited datasets.

Chapter 6. Summary of Recommendations for the Full-Scale Study

Methodologies and systems to accommodate the standard features of the Education Longitudinal Study of 2002 (ELS:2002) third follow-up—a mixed mode (web, computer-assisted telephone interview [CATI], and in the full-scale study, computer-assisted personal interview [CAPI]) survey of a dispersed youth sample—have been thoroughly tested in the prior study round, as well as in related longitudinal surveys of postsecondary populations, such as Beginning Postsecondary Students and Baccalaureate and Beyond. The function of the field test was more to confirm the efficacy of these standard locating practices and data collection methods than to break new ground that deviated from them.

Nonetheless, despite the familiar nature of the task, ELS:2002 third follow-up represents in many ways an extreme case of data collection difficulty, hence its own unique challenge. A sample of the modal age of 26 is highly mobile. Unlike any other current National Center for Education Statistics (NCES) study, there is a gap of a full 6 years between the second and third follow-up rounds of ELS:2002. Moreover, the ELS:2002 sample is not an "elite" sample such as baccalaureate recipients, but includes the full sociodemographic range of the young adult population, including difficult-to-survey groups such as high school dropouts. Panel maintenance activities (chiefly address updates) between the rounds do help to mitigate the negative effects of this long gap in the interview schedule, but cannot take the place of a full-fledged interview contact. A further challenge is to obtain an extremely high (90 percent or above) response rate.

Below, recommendations are provided for both pre-data collection (e.g., locating) and data collection procedures, and the general types of modifications that were applied in the questionnaire revision for the full-scale study are summarized. All recommendations that are made in this chapter will be implemented.

6.1 Recommendations for Pre-Data Collection Procedures

On the basis of interviewer performance in the field test, it is recommended that the same basic training protocols, and time allocations for training topics, be used in the full-scale study.

There were five locating methods used in the field test. All proved of some value, and given the targeted 90 percent response rate, it is recommended that all—batch tracing, mailings, e-mail follow-up, CATI locating and pre-intensive tracing, and intensive tracing—be implemented in the full-scale study.

6.2 Recommendations for Data Collection

In the full-scale study, additional effort will be needed to locate and interview high school dropouts and prior-round nonrespondents. The longer time frame of the full-scale data collection should be exploited to succeed with hard-to-locate populations. The full-scale study will also employ field interviewing and field locating, which will increase the resources to find and interview difficult-to-locate and difficult-to-contact sample members.

As done in the field test, emphasis should be placed on the role of parents as gatekeepers. Use of dual mailings, translation of contacting materials into languages other than English, and use of CATI procedures to elicit sample member addresses from parents, will all be important.

A variety of types of contacts and contacting materials, as used in the field test, are recommended for the full-scale study to ensure that sample members are both reached and reminded about the study. The contacts include a letter sent in a 9x12 envelope for the first contact, letters sent in standard-sized envelopes, e-mails, postcards, and text messages.

The propensity modeling experiment also holds important lessons for the full-scale study. Importantly, strategic case targeting can be implemented well on ELS:2002. Data on patterns of response and sample member characteristics supply a rich store of predictive variables for the third follow-up. In the experiment, interventions were effectively delivered to targeted cases providing confidence that this or a similar process could also be implemented successfully on the much larger study. However, this field test experiment revealed only small differences in measured bias; therefore, propensity to respond alone may not be the best method on which to base the targeting of cases. Although strategic case targeting is sensible and is consistent with recent developments in survey methodology, ELS:2002 has not yet identified the best statistical indicator on which to base strategic case targeting. New statistics, such as the Mahalanobis distance function, should be considered for the full-scale study in the context of survey designs with responsive and adaptive features.

6.3 Recommendations for the Study Instrument

On the basis of field test analysis of timing data, it is recommended that the questionnaire be slightly reduced in length. This is necessary to ensure an average interview duration of about 35 minutes.

Timing data were also used to identify lengthy items, the presentation of which might be revised in shorter form.

Although changes in specific items are too numerous to list, changes of several general types have been recommended: (1) revision or deletion based on item distributional properties such as lack of variance; (2) closing of response options for items with open “Please Specify” text; (3) revisions based on results of the reliability reinterviews; (4) revisions based on analysis of scale reliabilities; (5) changes made on the basis of methodological probes built into the field test questionnaire; and (6) changes based on findings from the two rounds of cognitive

interviews. Revisions and deletions were also informed by Technical Review Panel recommendations, and driven by reexamination of policy relevance or theoretical cogency of prime constructs, in deliberations between RTI and NCES staff.

6.4 Recommendations for the Survey Control Systems and Data Processing

Systems and processes used in the ELS:2002 third follow-up field test were designed and developed to test and identify areas of improvement in preparation for the full-scale study. However, no changes are in fact recommended at the systems level other than the addition of CAPI systems; the field test suggests that the full-scale study should be well served by the systems that were tested.

In data processing and delivery, one element—the creation of derived or composite variables—is normally omitted from the field test. With finalization of the full-scale questionnaire, it is recommended that identification and specification of composites take place as soon as possible, to give plenty of time to reflect on the variables chosen. Additionally, weights and imputations are not generated in the field test situation, but planning for these statistical procedures (as well as for disclosure avoidance) will be important full-scale activities.

This page left intentionally blank.

References

Cronbach, L.J. (1951). "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika*, 16: 297-334.

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.

Groves, R. M., & Heeringa, S. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A: Statistics in Society, 169*(Part 3), 439-457.

Ingels, S.J., Pratt, D.J., Wilson, D., Burns, L.J., Currivan, D., Rogers, J.E., and Hubbard-Bednasz, S. (2007). *Education Longitudinal Study of 2002 (ELS:2002) Base-Year to Second Follow-up Data File Documentation* (NCES 2008-347). U.S. Department of Education, Institute of Education Sciences. Washington, DC: National Center for Education Statistics.

Ingels, S.J., Pratt, D.J. Herget, D.R., Burns, L.J., Dever, J.A., Ottem, R., Rogers, J.E., Jin, Y., and Leinwand, S. (2011). *High School Longitudinal Study of 2009 (HSLS:09). Base-Year Data File Documentation*. (NCES 2011-328). U.S. Department of Education, Institute of Education Sciences. Washington, DC: National Center for Education Statistics.

Lent, R.W., Brown, S.D. & Hackett, C. (1994), "Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance." *Journal of Vocational Behaviour*, *45*, 79-122.

Nunally, J.C., and Bernstein, I.H. (1994). *Psychometric Theory*. 3rd ed. New York: McGraw-Hill.

Peytchev, A., Baxter, R. K., & Carley-Baxter, L. R. (2009). Not All Survey Effort is Equal: Reduction of Nonresponse Bias and Nonresponse Error. *Public Opinion Quarterly, 73*(4), 785-806.

Peytchev, A., Riley, S., Rosen, J.A., Murphy, J.J. and Lindblad, M. (2010). Reduction of Nonresponse Bias in Surveys through Case prioritization. *Survey Research Methods,* 4(1), 21-29.

Research Triangle Institute (2008). *SUDAAN Language Manual, Release 10.0.* Research Triangle Park, NC: Research Triangle Institute.

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.

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology, 35(1)*, 101-114.

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.

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.