Education Longitudinal Study of 2002 (ELS:2002)

Base-Year to Second Follow-up Data File Documentation

U.S. Department of Education
NCES 2008-347

October 2007

U.S. Department of Education
NCES 2008-347

Education Longitudinal Study of 2002 (ELS:2002)

Base-Year to Second Follow-up Data File Documentation

October 2007

Authors
Steven J. Ingels
Daniel J. Pratt
David Wilson
Laura J. Burns
Douglas Currivan
James E. Rogers
Sherry Hubbard-Bednasz
RTI International

John Wirt
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

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

October 2007

The NCES World Wide Web home page address is http://nces.ed.gov/.
The NCES World Wide Web electronic catalog address is http://nces.ed.gov/pubsearch/.

Suggested Citation

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

Contact:
John G. Wirt
(202) 502-7478
John.Wirt@ed.gov

Executive Summary

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

Chapter 2 describes the data collection instruments, including both the development and content of the tests and questionnaires used in the three rounds of data collection. It also documents the first follow-up transcript and course offerings studies and provides information about linkages to external data sources.

The sample design is documented in chapter 3, while data collection procedures and results are presented in chapter 4. Chapter 5 describes data preparation and processing, including data file preparation.

Chapter 6 provides an account of the weighting procedures used in the study, with special emphasis on the most recent (2006) round. The chapter also covers statistical procedures, such as imputation, disclosure avoidance, and the calculation of design effects. Chapter 7 describes the contents of the data files, including the data structure and analysis populations.

The appendixes include, among other topics, an introduction to the base-year to second follow-up electronic codebook (ECB); a flow chart and facsimile for the second follow-up instrument; a crosswalk between occupation coding schemes; a glossary of terms; information about making cross-cohort comparisons; a listing of the superset of variables to be found on the ELS:2002 second follow-up restricted-use ECB and the subset of the same variables provided by the ELS:2002 second follow-up Data Analysis System (DAS); a description of the second follow-up composite variables; and a synopsis of the ELS:2002 second follow-up field test.

Foreword

This manual has been produced to familiarize data users with the procedures followed for data collection and processing for the base year through second follow-up of the Education Longitudinal Study of 2002 (ELS:2002). It also provides the necessary documentation for use of the data files, as they appear on the ELS:2002 base-year to second follow-up electronic codebook (ECB) (NCES 2008-346), and information that may be helpful to users of the ELS:2002 Data Analysis System (DAS).

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

John Wirt
ELS:2002 Project Officer
Elementary/Secondary & Libraries Studies

Jeffrey Owings
Associate Commissioner
Elementary/Secondary & Libraries Studies

Acknowledgments

Jeffrey A. Owings served as the NCES project officer for the base year and first follow-up. John Wirt was the NCES project officer for the second follow-up. Daniel J. Pratt of RTI served as the Education Longitudinal Study of 2002 (ELS:2002) base-year through second follow-up project director. Steven J. Ingels of RTI was principal investigator.

Key RTI second follow-up task leaders were Robert Bozick, Laura Burns, Doug Currivan, Brian Evans, Saju Joshua, Tiffany L. Mattox, James Rogers, and David Wilson. Other RTI staff who played important roles in the ELS:2002 second follow-up were Chris Alexander, Brett Anderson, Donna Anderson, Kimberly Ault, Janet Austin, Edrina Bailey, Stephen Black, Ellen Causey, Jim Chromy, Elizabeth Copello, Lanting Dai, Ben Dalton, Marianne Daye, John Doherty, Kristin Dudley, David Higgins, Sherry Hubbard-Bednasz, Erich Lauff, Robert D. Lee, Ajay Maddi, Jeff Mahoney, Katherine Mason, Mani Medarametla, Chinh Nguyen, Joseph Nofziger, Bryce Norton, Jeremy Porter, Mike Planty (now of NCES), Neeraja Sathe, Ellen Scheib, Peter Siegel, Joe B. Simpson, Helen Smith, David Sroka, Milorad Stojanovic, Casey Tart, Chrystal Thompson, and Feng Yu. Many other individuals contributed to the success of the study, in their roles as interviewers, data collection supervisors, tracing specialists, or in data receipt and materials preparation.

We also wish to acknowledge Laura Horn and Lutz Berkner of our subcontractor MPR Associates, with whom we have collaborated in production of the ELS:2002 Data Analysis System.

The authors wish to thank the many individuals who assisted in the planning of the ELS:2002 second follow-up. We are particularly indebted to the ELS:2002 Technical Review Panel, whose members reviewed plans for the study, helped refine them, and provided important suggestions to help guide development of the instrumentation. The following individuals were active in the ELS:2002 Technical Review Panel in the second follow-up: Frank Balz, Tom Hoffer, Jacqueline King, Catherine Millett, Aaron Pallas, Kent Phillippe, and Leslie Scott.

Special thanks are in order to Jeffrey A. Owings, Associate Commissioner for Elementary/Secondary & Libraries Studies at NCES, in whose division ELS:2002 is housed, and to John G. Wirt, project officer for the second follow-up. Other NCES staff who have provided help, support, and assistance include Jim Griffith, Lisa Hudson, Andrew G. Malizio, Edith McArthur, Michael Ross, Neil Russell, and Marilyn M. Seastrom. Marsha Silverberg of the U.S. Department of Education and Clifford Adelman (formerly of the U.S. Department of Education) also contributed to the study.

Lastly, we want to acknowledge the help of four other individuals at RTI in the preparation of this document: Craig Hollingsworth and Michelle Myers, who edited the document, and Sharon Powell and Daliah Rainone, who provided document preparation and graphics support.

Contents

List of Tables

Table **Page**

List of Appendix Tables

Table **Page**

Table **Page**

Table **Page**

Table **Page**

Table **Page**

List of Figures

Chapter 1
Introduction

1.1 Overview of the Data File Documentation

This report provides guidance and documentation for users of the combined base-year through second follow-up data of the Education Longitudinal Study of 2002 (ELS:2002). ELS:2002 is sponsored by the National Center for Education Statistics (NCES) of the Institute of Education Sciences (IES), U.S. Department of Education. The base-year and follow-up studies were conducted through a contract to RTI International, a university-affiliated, nonprofit research organization based in North Carolina. This document contains information about the purposes of ELS:2002; the base-year, first, and second follow-up data collection instruments; the sample design; and the data collection and data processing procedures. The manual provides guidance for understanding and using data from all components of the base year and its two follow-ups.

The ELS:2002 base-year to second follow-up dataset has been produced in a restricted-use electronic codebook (ECB) version (NCES 2008-346) as well as a public-use web-only Data Analysis System (DAS). The data files reflect alteration or suppression of some of the original data. The data were edited to minimize the risk of disclosing the identity of responding schools and individuals. Although the primary focus of this manual is the ECB (because it is more inclusive), much of the information supplied is also applicable to the DAS version of the dataset. Because the ELS:2002 second follow-up ECB is restricted use only, second follow-up sample sizes in this report have been rounded to tens or hundreds (numbers of less than four digits have been rounded to tens; numbers of four or five digits have been rounded to hundreds). Because base-year and first follow-up data were earlier released on public-use ECBs, exact sample sizes—in conformity to previously released documentation and published reports—have been provided.

Chapter 1 addresses three main topics. First, it supplies an overview of the NCES education longitudinal studies program, thus situating ELS:2002 in the context of the earlier NCES high school cohorts studied in the 1970s, 1980s, and 1990s. Second, it introduces ELS:2002 by sketching some of the research and policy issues it can address and by delineating its study design. Third, it provides an overview of the various modes of data analysis that the design supports and touches on files and systems that have been provided for analysis.

In subsequent chapters, additional topics are addressed: instrumentation (chapter 2), sample design (chapter 3), data collection methods and results (chapter 4), data preparation and processing (chapter 5), weighting and estimation (including imputation, bias analysis, and design effect analysis) (chapter 6), and data file structure and contents (chapter 7).

Appendixes provide additional information, including special information on cross-cohort comparisons (appendix A), an introduction to the restricted-use ECB (appendix B), a synopsis of the ELS:2002 second follow-up field test (appendix C), base-year to first follow-up Data File Documentation errata (appendix D), flow chart and facsimile for the second follow-up questionnaire (appendix E), an occupational coding crosswalk (appendix F), transcript standard errors and design effects (appendix G), supplemental weighting nonresponse adjustment tables

(appendix H), average weight adjustment factors (appendix I), second follow-up design effects (appendix J), nonresponse bias tables (appendix K), documentation of differences between the public-use and restricted-use files (appendix L), a listing of all ECB and DAS variables (appendix M), further information about composite variables and ancillary or ecological data drawn from relevant extant databases (appendix N), and a glossary of terms (appendix O).

1.2 Historical Background

1.2.1 NCES High School Longitudinal Studies Program

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

NCES (and ELS:2002) 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 IES, in which NCES is now housed.

The high school longitudinal studies program consists of three completed studies: the National Longitudinal Study of the High School Class of 1972 (NLS:72), the High School and Beyond (HS&B) longitudinal study of 1980, and the National Education Longitudinal Study of 1988 (NELS:88). In addition, base-year through second follow-up data (2002–06) for ELS:2002, the fourth longitudinal study in the series, are now available. Taken together, these studies describe the educational experiences of students from 4 decades—the 1970s, 1980s, 1990s, and 2000s—and also provide bases for further understanding of the correlates of educational success in the United States. A fifth study, the High School Longitudinal Study of 2009 (HSLS:09) is presently in its design phase. Figure 1 is a temporal presentation of the four longitudinal high school cohort studies for which data are currently available, and highlights their component and comparison points. Figure 1 does not identify all future follow-up points for ELS:2002; final decisions have yet to be made concerning them. However, the general expectation is that the ELS:2002 cohorts will be followed until about age 26–30.

1.2.2 National Longitudinal Study of the High School Class of 1972

The National Education Longitudinal Studies program began over 30 years ago with the implementation of NLS:72.[1] NLS:72 was designed to provide longitudinal data for education policymakers and researchers who link educational experiences in high school with important downstream outcomes such as labor market experiences and postsecondary education enrollment and attainment. With a national probability sample of 19,001 high school seniors from 1,061

[1] For documentation on NLS:72, see Riccobono et al. (1981) and Tourangeau et al. (1987). While recent NCES reports and user documentation may be found on the NCES website (http://nces.ed.gov), some older documentation may be unavailable. NLS:72 and older HS&B manuals may be downloaded from the International Archive of Education Data at the Inter-university Consortium for Political and Social Research at the University of Michigan (http://www.icpsr.umich.edu). Materials may also be obtained in microfiche or photocopy format from the Education Resources Information Center database (http://www.eric.ed.gov).

public, Catholic, and other private schools, the NLS:72 sample was representative of approximately 3 million high school seniors enrolled in 17,000 U.S. high schools during the spring of the 1971–72 school year. Each member of this cohort was asked to complete a student questionnaire and a cognitive test battery. In addition, administrators at the sample members' schools were asked to supply information about the schools' programs, resources, and grading systems, as well as survey data on each student. No parent survey was conducted. However, postsecondary education transcripts were collected from the institutions attended by students. Five follow-up surveys were completed with this student cohort, with the final data collection taking place in 1986, when the sample members were 14 years removed from scheduled high school graduation and approximately 32 years old.

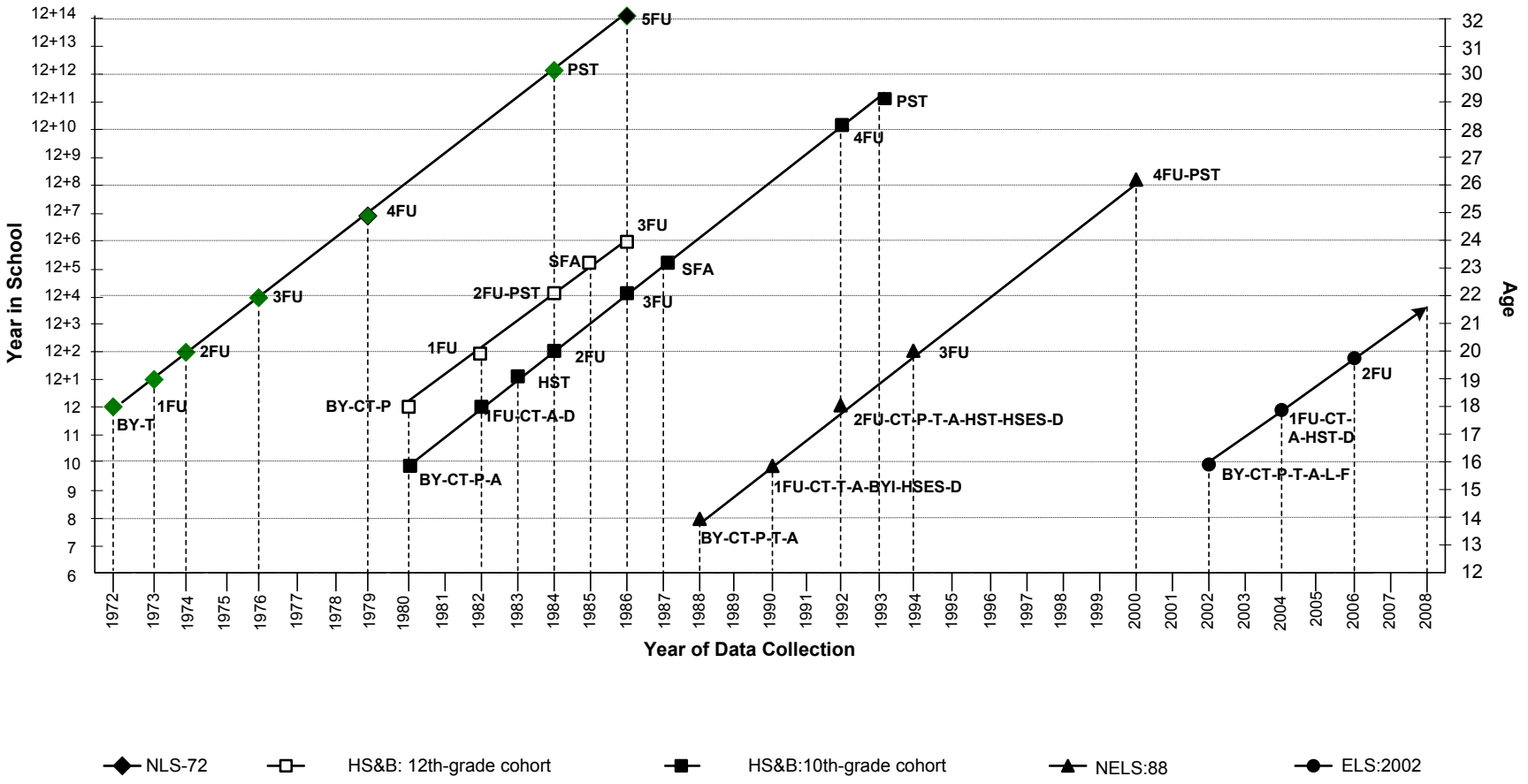
A wide variety of data was collected in the NLS:72 surveys. For example, in addition to background information about the students and their families, the base-year and follow-up surveys collected data on each respondent's educational activities (e.g., schools attended, grades received, and degree of 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.

1.2.3 High School and Beyond

The second in the series of NCES longitudinal studies was launched in 1980. HS&B included one cohort of high school seniors comparable to the NLS:72 sample; however, it also extended the age span and analytical range of NCES longitudinal studies by surveying a sample of high school sophomores. Base-year data collection took place in the spring term of the 1979–80 academic year with a two-stage probability sample. More than 1,000 schools served as the first-stage units, and 58,000 students within these schools were the second-stage units. Both cohorts of HS&B participants were resurveyed in 1982, 1984, and 1986; the sophomore group also was surveyed in 1992.[2] In addition, to better understand the school and home contexts for the sample members, data were collected from teachers (a teacher comment form in the base year asked for teacher perceptions of HS&B sample members), principals, and a subsample of parents. High school transcripts were collected for a subsample of sophomore cohort members. As in NLS:72, postsecondary transcripts were collected for both HS&B cohorts; however, the sophomore cohort transcripts cover a much longer time span (to 1993).

[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/.

Figure 1. Longitudinal design for the NCES high school cohorts: 2006



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

BY=Base-year data collection
1FU=1st follow-up data collection
2FU=2nd follow-up data collection
3FU=3rd follow-up data collection
4FU=4th follow-up data collection
5FU=5th follow-up data collection

CT=Cognitive test
P=Parent survey
T=Teacher survey
A=Administrator survey
L=Library/media center survey
F=Facilities checklist

HST=High School Transcript
PST=Post-Secondary Transcript
SFA=Student Financial Aid
BYI=Base-year Ineligible Study
HSES=HS Effectiveness Study
D=Dropout Survey

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002); National Education Longitudinal Study of 1988 (NELS:88); High School and Beyond Longitudinal Study (HS&B); and National Longitudinal Study of the High School Class of 1972 (NLS:72).

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 education reform movement that began in the early 1980s.[3]

1.2.4 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 premeasure of their achievement and status, prior to their entry into high school. NELS:88 represents an integrated system of data that tracked students from junior high or middle school through secondary and postsecondary education, labor market experiences, and marriage and family formation. Because ELS:2002 repeats so many of its innovations and design features, it will be useful to provide a detailed, round-by-round picture of NELS:88.

Data collection for NELS:88 was initiated with the 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 sample was also surveyed after scheduled high school graduation, in 1994 and 2000.[4]

1.2.4.1 NELS:88 Base Year

The NELS:88 base year (1988) successfully surveyed 24,599 students, out of some 26,432 selected 8th-graders, across 1,052 public, Catholic, and other private schools. In addition to filling out a questionnaire, students also completed assessments in four subjects (mathematics, science, reading, and social studies). The base year also surveyed one parent, two teachers, and the principal of each selected student. The base-year research instruments collected information about home, school, and individual factors that could serve as predictors for later outcomes (e.g., viewed in terms of positive outcomes, graduating from high school, making a smooth transition into the workforce, or completing postsecondary education). Information collected in the base year included family income, parental education, and occupation; parental aspirations for their 8th-grader; the 8th-grader's educational and occupational aspirations and plans, school

[3] For a summary of reforms instituted between the time the HS&B cohort was in high school and the NELS:88 cohort was in middle/junior high and high school, see Rasinski et al. (1993). For a summary of state education reforms instituted during the earlier school years of the ELS:2002 cohort, see Hurst et al. (2003).

[4] The entire compass of NELS:88, from its baseline through its final follow-up in 2000, is described in Curtin et al. (2002). Final outcomes for NELS:88 (in 2000) are reported in Ingels et al. (2002). The most extensive documentation of the NELS:88 assessment battery is found in Rock and Pollack (1995). The quality of NELS:88 data in the in-school rounds is examined in Kaufman and Rasinski (1991) and McLaughlin and Cohen (1997). The sample design is documented in Spencer et al. (1990). Eligibility and exclusion issues are addressed in Ingels (1996). 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 data (see http://nces.ed.gov/surveys/nels88/Bibliography.asp).

experiences, extracurricular activities, jobs and chores, television viewing, and reading; teacher perceptions of the 8th-grader's classroom performance and personal characteristics; curricular and instructional information about the classes in which teachers taught the 8th-grader; the teacher's own background and activities; and the principal's reports on the educational setting and environment of the school.

1.2.4.2 NELS:88 First Follow-up

A first follow-up took place in 1990. In the NELS:88 first follow-up (initial data release), there are 19,260 participants (18,220 students and 1,040 dropouts) from a sample of 20,700. (There were some changes to the file in the second follow-up rerelease of the 1990 data, which shows a revised sample size of 20,840.) The first follow-up sample was freshened to represent 1990 spring-term sophomores nationally. At that time, student cohort members, their teachers, and their principals were resurveyed. The first follow-up presented three major new analytic opportunities: (1) longitudinal analysis of gains in tested achievement and the correlates of achievement gains, (2) identification of high school dropouts and factors associated with persistence and dropping out, and (3) cross-cohort comparison (1990 high school sophomores could be compared to sophomores in 1980).

1.2.4.3 NELS:88 Second Follow-up

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

1.2.4.4 NELS:88 Third Follow-up

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

1.2.4.5 NELS:88 Fourth Follow-up

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

1.3 Education Longitudinal Study of 2002

ELS:2002 represents a major longitudinal effort designed to provide trend data about critical transitions experienced by students as they proceed through high school and into postsecondary education or their careers. The 2002 sophomore cohort is being followed, initially at 2-year intervals, to collect policy-relevant data about educational processes and outcomes. 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 surveyed and assessed in a national sample of high schools with 10th grades. Their parents, teachers, principals, and librarians were surveyed as well.

In the first of the follow-ups, base-year students who remained in their base-year schools were resurveyed and tested (in mathematics) 2 years later, along with a freshening sample that makes the study representative of spring-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. 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 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.

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

1.3.1 ELS:2002 Research and Policy Issues

Apart from helping to describe the status of high school students and their schools, ELS:2002 is providing information to help address a number of key policy and research questions. The study is intended to produce a comprehensive dataset for the development and evaluation of education policy at all government levels. Part of its aim is to inform decisionmakers, 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:

- students' academic growth in mathematics;
- the process of dropping out of high school;
- 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; such subgroups include 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 can be 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;
- the transition of those who do not go directly to postsecondary education or to the world of work; and
- access to and choice of postsecondary educational institutions.

Future data collections 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), seniors (2004), and post-sophomore-year dropouts (2004);
- longitudinal analysis (including examination of life course changes);
- cross-cohort comparisons with American high school students of earlier decades; 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).

1.3.2 ELS:2002 Study Design

ELS:2002 is designed to monitor the transition of a national sample of young people as they progress from 10th grade through high school and on to postsecondary education or the world of work, or both.

ELS:2002 has two distinctive features. First, it is a longitudinal study, in which the same units (schools and students) are surveyed repeatedly over time. Individual students have been followed through high school and will be followed for a number of years thereafter. The base-year schools were surveyed twice, in 2002 and in 2004. Second, in the high school years, ELS:2002 is an integrated, multilevel study that involves multiple respondent populations. The respondents include students, their parents, their teachers, and their schools (from which data are collected at four levels: from the principal, the librarian, a facilities checklist, and school course catalogues and records, which support a school course offerings component). Each of the two distinctive features—the longitudinal nature of the ELS:2002 design and its multilevel focus—will be explained in greater detail below.

The transition through high school and beyond into postsecondary institutions and the labor market is both complex (youth may follow many different paths) and prolonged (it takes place over a period of years). The complexity and time frame for this transition make longitudinal approaches especially appropriate. By surveying the same young people over time, it is possible to record the changes taking place in their lives. Gathering information about the ways that their earlier achievements, aspirations, and experiences predict what happens to the respondents later is also possible. In the baseline data collection (spring 2002), ELS:2002 measured students' tested achievement in reading and mathematics. ELS:2002 also obtained information from students about their attitudes and experiences.

These same students were resurveyed 2 years later (in 2004), in the ELS:2002 first follow-up, to measure changes such as achievement gains in mathematics and changes in enrollment status (e.g., the situation of students who drop out of school compared with those who persist in their education). The cohort members were resurveyed 4 years after the base year (2006), and the second follow-up data supply information about postsecondary educational access and choice, or transition to the labor market for cohort members who did not continue their education.

Cohort members will be followed for a number of years after this follow-up so that later outcomes (e.g., their persistence in higher education and baccalaureate attainment, or their success in the labor market) can be understood in terms of their earlier aspirations, achievement, and high school situation.

ELS:2002 gathers information at multiple levels. It obtains information not only from students and their school records, but also from students' parents, teachers, and the administrators (principal and library media center director) of their schools. Data from their teachers, for example, provide information both about the students' and the teachers' backgrounds and activities. This multilevel focus supplies researchers with a comprehensive picture of the home, community, and school environments and their influences on the student.

This multiple-respondent perspective is unified by the fact that for most purposes, the student is the basic unit of analysis.[5]

In addition, 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 and Private School Survey (PSS), as well as post-high school institutional information such as the NCES Integrated Postsecondary Education Data System. Additional sources that have been drawn upon or linked to include student application and loan information, including the Free Application for Federal Student Aid, and various sources of test scores (SAT, ACT, and the GED testing program) and the National Student Loan Data System.

With the addition of postsecondary data in the 2006 second follow-up, ELS:2002 greatly enlarges its ability to connect high school antecedents to later outcomes. For students who continue on to higher education, researchers can use ELS:2002 to measure the effects of their high school careers on subsequent access to postsecondary institutions, their choices of institutions and programs, and, as time goes on, their postsecondary persistence, attainment, and eventual entry into the labor force and adult roles. For students who go directly into the 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.

1.3.2.1 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[6] participating schools and 15,362 participating students. Schools are the first-stage unit of selection, with sophomores randomly selected within schools.

[5] Base-year school administrator, library media center, and facilities data can be used to report on the nation's schools with 10th grades in the 2001–02 school year. A first follow-up course offerings file further enriches the information available about high schools with 10th grades in 2002. However, if history is a guide, most analysts will employ the school-level data to provide further contextual information on the student.

[6] Note that exact sample sizes are provided for the base year and first follow-up of ELS:2002, consistent with past documentation (NCES 2004-405, NCES 2006-344) and the released public-use files in ECB format. However, since there is no public release file for the second follow-up, exact sample sizes are not given for the 2006 round. Rather, to perturb the data, as is required in reporting on restricted-use files, sample sizes of less than four digits are rounded to tens, and sample sizes of four or five digits are rounded to hundreds.

- Oversampled Asian[7] and Hispanic students and private schools.
- Designed linkages with PISA (reading in 2000 and math in 2003) and National Assessment of Educational Progress (NAEP 2005 math); 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 can be measured and related to school processes and mathematics coursetaking. The student questionnaire gathered information about the student's background, school experiences and activities, plans and goals for the future, employment and out-of-school experiences, language background, and psychological orientation toward learning.

One parent of each participating sophomore was asked to respond to a parent survey. The parent questionnaire was designed to gauge parental aspirations for their child, home background and the home education support system, the child's educational history prior to 10th grade, and parental interactions with and opinions about the student's school. For each student enrolled in English or mathematics, a teacher was also selected to participate in a teacher survey. The teacher questionnaire collected the teacher's evaluations of the student and provided information about the teacher's background and activities. The head librarian or media center director at each school was asked to complete a library media center questionnaire, which inquired into the school's library media center facility, its staffing, its technological resources, collection and expenditures, and scheduling and transactions. Finally, the facilities checklist was a brief observational form completed for each school. The form collected information about the condition of school buildings and facilities.

1.3.2.2 First Follow-up (2004)

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

- Most sample members were seniors, but some were dropouts or in other grades (early graduates or retained in an earlier grade).
- 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), dropout questionnaire, assessment in mathematics, and school administrator questionnaire were administered.

[7] 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.

- 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.
- 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 as well as 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 owing to severe disability or insufficient command of the English language such that they could not complete a questionnaire; and
- students at the 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.

While 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 that all sample members would respond to, 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). In chapter 2, the

various questionnaires—student, abbreviated student, transfer student, early graduate, homeschool, out-of-school (dropout), and new student supplement—are briefly described.

For some classifications of the sample, a first follow-up test score in mathematics was either collected (students still in the base-year school) or imputed (students who have transferred to a new school). For other categories of sample members, such as dropouts, early graduates, and homeschooled students, a test score has neither been collected nor imputed. (Note that missing base-year test score data have been imputed for base-year nonrespondents who became respondents in the first follow-up.)

For all classifications of sample members, information about student coursetaking (covering all years of high school and including the sequence in which courses were taken and grades earned) were collected late in 2004 and early 2005 through the high school transcript component of the ELS:2002 first follow-up study. Further information about the transcript component is contained in this volume and in Bozick et al. (2006).

At the school level, the first follow-up extended information about base-year schools through administration of a school administrator questionnaire. In addition, information about school course offerings was collected in the first follow-up transcript study. Finally, further information about participating schools at the time of the first follow-up survey can be obtained on the restricted-use ECBs by linking (via the NCES identification code [NCESID]) to the CCD or PSS, and, via ZIP codes, to 2000 Census data. The NCES school district database and its Census data also are accessible on the restricted-use file by means of the NCESID.

1.3.2.3 Second Follow-up (2006)

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

- Post-high-school follow-up with web-based instrument for self-administration, computer-assisted telephone interview (CATI), or computer-assisted personal interview (CAPI).
- Survey 2 years after the cohorts' modal high school graduation captures six distinct groups:
  - high school late completers;
  - nonenrollers in higher education;
  - prompt postsecondary education enrollers;
  - delayed postsecondary education enrollers;
  - higher education leavers (versus persisters) and returnees; and
  - delayer-leavers.
- Three distinct (and sometimes alternating or combined) transitions:
  - transition to the work force;
  - transition 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[8] cases, about 14,200 sample members completed interviews, for a weighted response rate of 88 percent.

The ELS:2002 second follow-up provides data to map and understand a key transition: the transition of the majority of cohort members out of high school. For the cohort as a whole, the second follow-up obtained information that will permit researchers and policymakers to better understand issues of postsecondary educational access and choice. Thus, a major focus of the second follow-up interview 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 (or family formation). 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, community involvement, and negative life events.

The principal innovation of the ELS:2002 second follow-up—one that represents a technological improvement over the data collection methods used in its predecessor, NELS:88—is application of computer methods to self-administered questionnaires for the out-of-high-school population, in which the questionnaire is completed on the Web. The survey used a web-enabled survey system to program the instrument for self-administration. The same electronic instrument was used in the CATI and CAPI instruments as well. (The self- and interviewer-administered survey instruments are indistinguishable in terms of screen text and skip patterns in each of the three modes.) The advantages of a web-based instrument include real-time data capture and access, including data editing in parallel with data collection.

1.3.2.4 Further Follow-ups

The number of (and dates for) further web/CATI/CAPI and postsecondary education transcript follow-ups will be determined at a later date.

1.3.3 ELS:2002 Modes of Data Analysis

1.3.3.1 Cross-Sectional Profiles

Cross-sectional data permit characterization of the nation's high school sophomores in the spring term of the 2001–02 school year. Initial cross-sectional findings from the base year are

[8] As earlier noted, only approximate sample sizes are provided for the 2006 round, because restricted-use data are used. Exact sample sizes from restricted-use data cannot be published unless the data are perturbed in some ways. The perturbation approach taken here was to round the exact sample sizes to tens (for one- to three-digit numbers) or hundreds (for four- to five-digit numbers). In contrast, a public-use ECB was produced for the 2002 (base-year) and 2004 (first follow-up) rounds. For this reason, exact sample sizes can be reported for the earlier rounds.

available in an NCES report, *A Profile of the American High School Sophomore in 2002*.[9] Because of sample freshening, the results 2 years later provided a basis for profiling the nation's high school seniors in the spring term of the 2003–04 school year. A report on seniors has also been released[10] as well as findings pertaining to high school graduates that uses data from the ELS:2002 high school transcript study.[11] Finally, a "first look" report containing some basic tabulations of second follow-up data accompanies the release of the 2002–06 combined data.

1.3.3.2 Longitudinal Analysis

Longitudinal analysis became possible with the release of data from the 2004 first follow-up[12] and has been further extended by the addition of the 2006 data point. The primary research objectives of ELS:2002 are longitudinal in nature. The study provides the basis for within-cohort comparison by following the same individuals over time to measure postsecondary educational and workforce entry and relate these outcomes to antecedents identified in earlier rounds, including individual, home, school, and community factors.

1.3.3.3 Cross-cohort Comparisons

As part of an important historical series of studies that repeats a core of key items each decade, ELS:2002 offers the opportunity for the analysis of trends in areas of fundamental importance, such as patterns of coursetaking, rates of participation in extracurricular activities, academic performance, and changes in goals and aspirations. An NCES report is available that details the experiences of HS&B, NELS:88, and ELS:2002 high school sophomores.[13] With completion of the second follow-up in 2006, researchers can now compare ELS:2002 high school seniors' experiences 2 years out of high school with those of the NELS:88 cohort in 1994, HS&B in 1982 and 1984, and NLS:72 in 1974. With the ELS:2002 academic transcript data, researchers can also make trend comparisons with academic transcript data containing students' high school course histories and sequences because comparable transcript studies have been conducted with spring-defined senior cohorts, starting with HS&B[14] (1982) and including NELS:88 (1992) and NAEP (1987, 1990, 1994, 1998, 2000, and 2005). (See appendix A.)

1.3.3.4 International Comparisons

A feature of ELS:2002 that expands the study's power beyond that of the predecessor studies is that it can be used to support international comparisons. A concordance has been generated to link the ELS:2002 scale to that of PISA. The Organization for Economic Cooperation and Development's PISA (Lemke et al. 2001, 2004) is an internationally

[9] See Ingels et al. (2005a) (NCES 2005-338). A small, but growing, ELS:2002 bibliography can be found at http://nces.ed.gov/surveys/els2002/Bibliography.asp.

[10] See Ingels, Planty, and Bozick (2005), *A Profile of the American High School Senior in 2004* (NCES 2006-348).

[11] See Planty, Bozick, and Ingels (2006), *Academic Pathways, Preparation, and Performance—A Descriptive Overview of the Transcripts from the High School Graduating Class of 2003-04* (NCES 2007-316).

[12] For an example of longitudinal analysis, see Bozick and Ingels (2007), *Mathematics Coursetaking and Achievement at the End of High School: Evidence from the Education Longitudinal Study of 2002* (NCES 2007); or Bozick and Lauff (2007), *A First Look at the Initial Postsecondary Experiences of the Sophomore Class of 2002* (ELS:2002) (NCES 2008-308).

[13] See Cahalan et al. (2006), *United States High School Sophomores: A Twenty-Two Year Comparison, 1980–2002* (NCES 2006-327). A cross-cohort analysis of coursetaking trends, based on academic transcripts, has also been completed—see Dalton et al. (2007), *Advanced Mathematics and Science Coursetaking in the Spring High School Senior Classes of 1982, 1992, and 2004* (NCES 2007-312).

[14] However, the HS&B sophomore cohort 2 years later (1982) did not include a freshening sample of seniors; this introduces a small conservative bias in its estimates (see Dalton et al. 2007 for details).

standardized assessment administered to 15-year-olds in groups in their schools. PISA covers three domains: reading literacy, numeracy, and scientific literacy; ELS:2002 test results have been linked to PISA reading (2000) and mathematics (2003) scores so that the PISA scale can be used in ELS:2002 analyses. PISA aims to define each domain not merely in terms of mastery of the school curriculum, but also in terms of important knowledge and skills needed in adult life. Emphasis is placed on the mastery of processes, the understanding of concepts, and the ability to function in various situations within each domain.

1.3.4 Analysis Files and Systems

While the base-year and base-year to first follow-up ELS:2002 data deliveries include both a public-use ECB and a restricted-use ECB, there is only a restricted-use ECB for the combined base year to second follow-up. Restricted files require that analysts obtain a special institutionally based license from NCES. However, a base-year to second follow-up web-housed public-use DAS has also been produced. Full details about the ECB are provided in later chapters, particularly chapter 7. A "quick guide" to use of the base-year to second follow-up ECB appears as appendix B of this document. Although this document is primarily oriented to the ECB, information that will be helpful to DAS users is also included.

1.4 High School Longitudinal Study of 2002

A fifth study in the series—the High School Longitudinal Study of 2009 (HSLS:09)—is currently in its development phase. HSLS:09 will survey a nationally representative sample of high school students, their parents, teachers, and school administrators at several time points during students' secondary and postsecondary years. In the high school years, it will include assessments in both mathematics and science. Unlike previous studies in the series, HSLS:09 will collect data from students in the fall of their 9th-grade year, with a second round of data collection at the end of 11th grade in 2012, when most of the student cohort will be completing their junior year. The new schedule will allow researchers and policymakers to learn if and how 9th-grade plans are linked to students' subsequent behavior, from coursetaking to postsecondary choices, and how these plans evolve over time. In subsequent waves of data collection, the sample members will be followed into college and beyond.

Chapter 2
Base-Year Through Second Follow-up Instrumentation

2.1 Introduction

This chapter is divided into five main sections. Section 2.1 is an introduction to instrumentation issues. Section 2.2 provides information about the base-year and first follow-up questionnaires. Section 2.3 describes the base-year and first follow-up achievement tests. Section 2.4 introduces the academic transcript component. Finally, section 2.5 provides information about the ELS:2002 second follow-up (2006) questionnaire.

The *base-year* (2002) data collection instruments for the Education Longitudinal Study of 2002 (ELS:2002) consisted of five separate questionnaires (student, parent, teacher, school administrator, and library media center), two achievement tests (assessments in reading and mathematics), and a school observation form (facilities checklist).

The *first follow-up* (2004) data collection instruments comprised seven questionnaires and an achievement test in mathematics. The first follow-up questionnaires included a student questionnaire, a transfer student questionnaire, a new participant student questionnaire (NPSQ), a homeschool student questionnaire, an early graduate questionnaire, a dropout (not currently in school) questionnaire, and a school administrator questionnaire. A new participant supplement (NPS) (repeating questions from the base year) and an abbreviated version of the student questionnaire were also offered.[15] The base-year and first follow-up questionnaires can be found as portable document format (PDF) files on the National Center for Education Statistics (NCES) ELS:2002 website (http://nces.ed.gov/surveys/els2002/).

In the first follow-up, information was also collected about the course offerings of the base-year schools, as well as the transcript records (including both courses taken and grades and credits received) of the sophomore and senior cohorts.

In the *second follow-up* (2006), a single electronic questionnaire was administered in three modalities—a web-enabled self-administration, an interviewer administration of computer-assisted telephone interviews (CATI), and computer-assisted personal interviews.

2.1.1 Instrument Development Process and Procedures

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

1. *Sharing of draft data elements*. Draft elements of the questionnaires were shared with other government agencies, policy groups, and interested parties.
2. *Technical review panel (TRP) review*. The ELS:2002 TRP, a specially appointed, independent group of substantive, methodological, and technical experts, reviewed the questionnaires.
3. *NCES review*. The questionnaires underwent interdivisional review at NCES.

[15] In fact, the new participant student questionnaire is simply the new participant supplement and abbreviated first follow-up student questionnaire, joined together to create one booklet, for convenience of administration.

4. *Questionnaire revision*. The survey instruments were revised based on reviewer comments.
5. *Writing of justification*. A justification was written for the data elements, noting issue areas, constructs to be measured within each, and items that would be used to measure each construct.
6. *Office of Management and Budget (OMB) review*. The federal OMB reviewed the instruments.
7. *Questionnaire revision*. The questionnaires were revised based on OMB comments.
8. *Field testing and revision*. The instruments were field tested and revised based on field test results.

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

The field testing of procedures, questionnaires, and assessments was an especially important step in the development of the full-scale surveys. Field test instruments were evaluated in a number of ways. For the questionnaires, field test analyses included evaluation of item nonresponse, examination of test-retest reliabilities, calculation of scale reliabilities, and examination of correlations between theoretically related measures. For the achievement tests in mathematics and reading, item parameters were estimated for both 10th and 12th grade in the base-year field test. Both classical and Item Response Theory (IRT) techniques were employed to determine the most appropriate items for inclusion in the final (base-year main study) forms of the two tests. Psychometric analyses included various measures of item difficulty and discrimination, investigation of reliability and factor structure, and analysis of differential item functioning. In the first follow-up field test, similar classical and IRT psychometric analyses were conducted but with a slightly different end in terms of final format: adaptiveness was ensured through a two-stage test in the base year, whereas the test designed for the first follow-up main study based assignment of form on the base-year mathematics ability estimate. The base-year field test report is available from NCES (Burns et al. 2003). Findings of the first follow-up field test are summarized in appendix J of Ingels et al. (2005b), while second follow-up field test results are reported in appendix C of this volume.

2.1.2 Instrument Development Goals and Constraints

The primary research objectives of ELS:2002 are longitudinal in nature; therefore, the first priority was to select the items that would prove most useful in predicting outcomes as measured in future survey waves.

The second priority was to obtain needed cross-sectional data, whenever consistent with the longitudinal objectives, particularly data that could be used for cross-cohort comparison with past studies or linkage to certain current data collection efforts. Wherever possible, all ELS:2002 instruments were designed to provide continuity and consistency with the earlier education longitudinal studies of high school cohorts. Where appropriate, ELS:2002 drew items from the National Longitudinal Study of the High School Class of 1972, the High School and Beyond (HS&B) longitudinal study, and, most particularly, the National Education Longitudinal Study of

1988 (NELS:88). In addition, the study used coding frames and taxonomies that were comparable to those employed in past high school transcript studies, or (in the case of occupation coding) could be crosswalked to them. Apart from the cross-cohort comparisons that can be sustained through use of the test, questionnaire, and transcript data, ELS:2002 provides score linkages with the testing programs of the Program for International Student Assessment (PISA) (reading and mathematics) and National Assessment of Educational Progress (NAEP) (mathematics).

Although maintaining trend items to support cross-cohort comparisons was a major aim of instrument development, there was also a need to provide new items to address new areas of policy concern and to reflect recent advances in theory. For example, in the base year in particular, educational technology items were developed to reflect the fact that computers have become a major factor in learning in recent years. Psychological scales that reflect recent work in self-efficacy theory and related areas were also added.

Another consideration in the development of the ELS:2002 instruments was the need to obtain factual information from the best source among the various respondent populations. This was an issue both for the base year, in which both parents and students were surveyed, and first follow-up, where administrative records were pursued (transcript component) as well as self-reports (questionnaire).

2.2 Base-Year and First Follow-up Questionnaires

2.2.1 Base-Year Questionnaires

The various ELS:2002 base-year questionnaires can be found at http://nces.ed.gov/surveys/els2002/index.asp. Some detail about them is provided below.

2.2.1.1 Student Questionnaire

The ELS:2002 base-year student questionnaire was typically self-administered. Sophomore sample members normally completed the questionnaire in a group setting in their schools. A small number of students were surveyed outside of school, with a shortened version of the questionnaire in a CATI. Assessments in reading and mathematics were given at the same time (i.e., during the group administration), in a two-stage process in which the first stage was a routing test. The full questionnaire was available only in English, although a shortened Spanish version was also produced.

The student questionnaire was divided into seven sections: (1) locating information, (2) school experiences and activities, (3) plans for the future, (4) non-English language use, (5) money and work, (6) family, and (7) beliefs and opinions about self.

2.2.1.2 Parent Questionnaire

The parent questionnaire was to be completed by the parent or guardian most familiar with the sophomore's school situation and experience. Guided by this definition of the preferred respondent, the parent survey respondent was self-selected.

The parent questionnaire was available in both English and Spanish. Both a hardcopy version and an electronic CATI version[16] were produced. The parent questionnaire addressed the following five topic areas: (1) family background, (2) their child's school life, (3) their child's family life, (4) their opinions about their child's school, and (5) their aspirations and plans for their child's future.

2.2.1.3 Teacher Questionnaire

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

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

2.2.1.4 School Administrator Questionnaire

The base-year school administrator questionnaire collected information on the school in six areas: (1) school characteristics, (2) student characteristics, (3) teaching staff characteristics, (4) school policies and programs, (5) technology, and (6) school governance and climate. The school administrator data can be used contextually, as an extension of the student data, when the student is the fundamental unit of analysis. At the same time, the ELS:2002 base-year school sample is nationally representative and can stand alone as a basis for generalizing to the nation's regular high schools with sophomores in the 2001–02 school year.

2.2.1.5 Library Media Center Questionnaire

For the school library media center component, the school librarian, media center director, or school administrator supplied information about library media center size, organization, and staffing; technology resources and electronic services; extent of library and media holdings, including both collections and expenditures; and levels of facility utilization, including scheduling for use by students and teachers. Finally, the questionnaire also supplied information about the library media center's use in supporting the school's curriculum; that is, how library media center staff collaborate with and support teachers to help them plan and deliver instruction. Information in the library media center questionnaire can be used as contextual data with the student as the unit of analysis or to generalize to libraries within all

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

regular high schools with 10th grades in the United States in the 2001–02 school year (for ELS:2002 library component findings, see Scott 2004).

2.2.1.6 School Facilities Checklist

Instrumentation for the facilities component comprised a checklist to be completed by the survey administrator. The survey administrator was asked to observe a number of conditions at the school, including the condition of the hallways, main entrance, lavatories, classrooms, parking lots, and surrounding neighborhood. Of special interest were indicators of security (metal detectors, fire alarms, exterior lights, fencing, security cameras, etc.) and maintenance and order (trash, graffiti, clean walls and floors, noise level, degree of loitering, etc.). Information gathered in the facilities checklist can be used as contextual data with the student as the unit of analysis, or data can be used at the school level to generalize to all regular high schools with 10th grades in the United States in the 2001–02 school year. (For findings drawing on the Facilities Checklist, see Planty and DeVoe 2005.)

2.2.2 First Follow-up Questionnaires

The various ELS:2002 first follow-up (2004) questionnaires can be found at http://nces.ed.gov/surveys/els2002/index.asp. Some detail about them is provided below.

2.2.2.1 Introduction

The following questionnaires were employed in the ELS:2002 first follow-up: student questionnaire, dropout questionnaire, early graduate questionnaire, transfer student questionnaire, homeschool student questionnaire, and NPSQ. A school administrator questionnaire was also offered. For the ELS:2002 data user, it is necessary to specify which items are common to various questionnaires and which are unique, and how each questionnaire group relates to the analytic populations of interest. The ELS:2002 *Base-Year to First Follow-up Data File Documentation* (Ingels et al. 2005b, NCES 2006-344) includes as its table 2 a crosswalk that shows shared and unique items across the first follow-up questionnaires.

2.2.2.2 Questionnaire Assignment and Content

First follow-up student questionnaire assignment and content. The student questionnaire was administered to sophomore cohort members who had remained in their base-year school as well as to a freshening sample of spring-term 12th-graders in those same schools. Students who completed the student questionnaire also were normally eligible for the first follow-up mathematics assessment. Some students were administered an abbreviated version of the questionnaire (these cases are flagged on the data file). The questionnaire was primarily self-administered in in-school survey sessions, and secondarily, for some students, out of school through CATI or occasionally through mail or field interviews.

Some alterations were required to adapt the paper-and-pencil questionnaire to CATI. Generally, the wording of the paper-and-pencil questions was made more conversational for the telephone interview to facilitate interviewer-respondent interaction. On occasion, adaptations were made to account for the fact that those interviewed by telephone did not have the benefit of seeing the entire question with all of its elements at once. For example, students were asked to report how much coursework they had taken in various subject areas. Respondents who completed the paper-and-pencil form were able to see the full range of mathematics courses

listed more or less in the sequence in which they are taught. In this context, it was clear to respondents that "general math" referred to a basic math course as opposed to a catchall category. However, without the visual cues, telephone respondents may have misinterpreted general math to include all math courses. Therefore, for the telephone interview, general math was moved to the end of the list of math courses. Similar adaptations were required for the other telephone-administered questionnaires as well (transfer student, dropout, and so on). Generally, CATI telephone data collection took place subsequent to in-school data collection. Also, there was more ambiguity about the status (dropout, early graduate, transfer, homeschooled, and so on) of sample members interviewed outside the school setting. For this reason, the CATI interview included a series of screening questions to ensure that the proper questionnaire was administered. Such a screener was also used for field cases subject to in-person interview.

The student questionnaire comprised eight content modules. *Part I* of the questionnaire requested contact information in support of the longitudinal design.

Part II covered the student's school experiences and activities. Data generated from this section provide information about extracurricular participation, computer use in English and math, the transition process from sophomore year to upper-level secondary school, and the relationship of curricular programs and coursetaking to educational achievement and persistence. Some of these data may be viewed as outcomes, influenced by factors studied in the base year, and others as predictors of outcomes in future rounds.

Part III, "How You Spend Your Time," inquired about time usage on homework, television viewing, video and computer games, computers, nonschool reading, library use, and other activities. *Part IV* focused on plans and expectations for the future. It included questions that elicited information about students' educational and life goals and values. *Part V*, on education after high school, contained items on postsecondary planning steps and choice criteria. *Part VI* dealt with plans for work after high school. *Part VII* inquired about working for pay, including hours worked per week. Finally, *Part VIII* consisted of items on community, family, and friends.

First follow-up dropout questionnaire assignment and content. Dropouts were defined as sophomore cohort members who were out of school in the spring term of 2004, who had not received a high school diploma or General Educational Development (GED) credentials on or before March 15, 2004, and who had missed 4 or more consecutive weeks not due to accident or illness. Students who had a dropout episode but who had been in school for at least 2 weeks at the time of their school's Survey Day were administered the student questionnaire. The dropout questionnaire was administered in multiple modalities—self-administration, in-person interviewer administration, and over the telephone by means of CATI.

There was considerable overlap between the student and dropout questionnaires. *Part I* collected locating information for longitudinal follow-up. *Part II* contained items on school experiences and activities. Dropouts were asked questions about the school they last attended and their participation in alternative educational programs. In addition, they were asked to supply their specific reasons for leaving school prior to graduation. They were asked as well about plans to get a GED or return to high school. *Part III* covered time use (reading, library patronage, television, videogames, computer use, and so on). *Part IV* asked about plans and expectations for the future. *Part V* provided information to identify the type and amount of work that dropouts were engaged in. It gathered information about students' work status and history, how much they

earned, and how many hours they worked. *Part VI* asked about volunteer or community service work and the educational behaviors of friends.

Early graduate questionnaire assignment and content. Early graduates were interviewed outside the school setting, in multiple data collection modalities but most commonly by telephone. Early graduates were defined as sophomore cohort members who had graduated from high school or received a GED on or before March 15, 2004. The approach to early graduates differs somewhat across the several NCES high school cohort studies. In HS&B, the group that was captured was high school completers who finished early (i.e., prior to March 1, 1982). In NELS:88 and ELS:2002, an additional group is included, those who completed by alternative means (e.g., GED) prior to their classmates who were in the modal graduation sequence. In both HS&B and NELS:88, early graduates completed supplementary questions in addition to the full student questionnaire (answering from the vantage point of their recent high school experience). In ELS:2002, early graduates completed only a subset of the items on the student questionnaire, complemented by additional items pertaining to their situation. More specifically, early graduates were asked with whom they consulted when deciding to graduate early, the basis for that decision, and the means by which they did so. They also provided a history of their work and educational experiences since leaving high school.

Transfer student questionnaire assignment and content. Sophomore cohort members who had transferred out of their base-year school to a new school received the transfer student questionnaire. Transfer students were asked a subset of items from the student questionnaire, covering the following topics: school experiences and activities; time use; plans and expectations for the future; education after high school; work after high school; and community, family, and friends. In addition, transfer students were asked when they transferred and their reasons for doing so. Transfer students did not complete a cognitive test, but their test scores have been imputed. Thus, 2004 math scores are available for both sophomore cohort "movers" and "stayers" as well as freshened spring seniors (though not for dropouts or the homeschooled).

Homeschool student questionnaire assignment and content. ELS:2002 does not provide a representative sample of homeschooled high school students. (In the base year, all study sophomores were selected from regular U.S. high schools.) Instead, homeschooled students in ELS:2002 generalize only to sophomores in regular high schools in spring term 2002 who were in a homeschool situation 2 years later. The primary motive for administering a separate questionnaire to this subset of the sophomore cohort was that neither the transfer student questionnaire items nor the dropout items fully fit their situation.

Homeschooled students were asked about their schooling activities and status, including their grade, coursework completed in science and math, and steps taken toward college; how they spend their time; their plans and expectations for the future, including education and work after high school; work experiences; and community, family, and friends.

New participant supplement questionnaire assignment and content; NPS. There are essentially three categories of students who were ELS:2002 new participants in the first follow-up. One class is the spring-term high school seniors who entered the study through the freshening sample. A second class of new participants is that of base-year nonrespondents who completed a questionnaire in the first follow-up. The third and final class is that of sophomore cohort

members who were questionnaire-incapable[17] in 2002 because of disability or a language barrier, but who were reclassified as capable of completing a questionnaire in 2004. (An example might be an English language learner who was not proficient in English in 2002 but, with 2 additional years of instruction, had reached a level of English proficiency sufficient to deal with the ELS:2002 first follow-up questionnaire.) While the first of these three classes is by definition a student, the second and third groups include both students and out-of-school members of the sophomore cohort (such as dropouts and early graduates).

Any student new to the study at any of the core (base-year) schools was administered the NPSQ. However, transfer students and out-of-school cohort members were administered the relevant questionnaire and an NPS containing the key base-year items. For example, any student new to the study who had transferred to a new school was administered the transfer student questionnaire and an NPS. Any new respondent who was out of school, however, such as a dropout or early graduate, was administered the appropriate out-of-school questionnaire, as well as an NPS. Table 1 summarizes, for all new participants, use of the NPS and NPSQ, as well as base-year and first follow-up assessment status.

Table 1. Base-year key variables and test data available, by type of first follow-up new participants: 2004

First follow-up new participants	Source of base-year standard classification variables	Availability of base-year reading and math scores	Availability of first follow-up math scores
Sophomore cohort members in core (base-year) schools in 2004	NPSQ	Imputed	Tested
Sophomore cohort members in new schools in 2004	NPS	Imputed	Imputed
Sophomore cohort members out of school in 2004: dropouts	NPS	Imputed	—
Sophomore cohort members out of school in 2004: early graduates	NPS	Imputed	—
Freshened spring 2004 seniors	NPSQ	—	Tested
Sophomore cohort members homeschooled in 2004	NPS	Imputed	—

— Not available.

NOTE: NPSQ = New Participant Supplement Questionnaire. NPS = New Participant Supplement; this instrument contains only the key base-year items.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

The NPSQ gathered information that had been collected (for other students) in the base year on new participants' demographic characteristics, parental education and occupation, and language use. These items are identical to those on the NPS. In addition, a subset of items included on the student questionnaire was also posed to new participants. These items (which are identical in content to the abbreviated student questionnaire) relate to topics such as school experiences and activities; time use; plans and expectations for the future; education and work

[17] Students who were questionnaire-incapable were ineligible for the assessment and were ineligible for the questionnaire, based on language barriers or severe disabilities. Nonetheless, contextual data were gathered for them in the base year, and in the first follow-up, transcripts were collected and their questionnaire status was re-assessed, in order to capture any change in status. In some ELS:2002 documentation, the questionnaire-incapable group is referred to as "questionnaire-ineligible."

after high school; and work, community, family, and friendship experiences. In contrast, the NPS gathered the key base-year variables that also were included on the NPSQ.

School administrator questionnaire content and content linkages. The school administrator questionnaire collected information on the school in four areas: school characteristics, structure, and policies; student characteristics and programs; teacher and library staff characteristics; and principal reports on the school environment. Many school-level variables of analytic interest also pose a high risk of disclosure of school identities. For this reason, a number of analysis variables have been limited to the restricted-use electronic codebook (ECB) or may be accessed through a link provided only on the restricted-use ECB.[18]

School-level data are not nationally representative of American high schools in 2004, because the first follow-up sample did not factor in "births" of new schools and "deaths" of existing schools between 2002 and 2004. First follow-up school data, however, do provide a statistical portrait of a nationally representative sample of American high schools with 10th grades in 2002 (2 years later).

2.3 ELS:2002 Base-Year and First Follow-up Assessment Battery

Before considering test development and the mathematics and reading assessment frameworks, it is useful, as a point of entry into the first follow-up achievement tests, to consider the fact of test availability in conjunction with the main sample populations for which questionnaires were designed. As table 2 makes clear, the entire responding questionnaire-capable sophomore cohort was eligible to be tested in the base year. However, as table 2 also makes clear, not all groups were tested in the first follow-up, nor were test scores imputed for all groups.

Table 2. Assessment availability status, by sample group: 2004

Sample group (status in 2004)	Base year	First follow-up
2002 sophomores in core (base-year) schools in 2004	Tested[1]	Tested[2]
2002 sophomores in transfer schools in 2004	Tested[1]	Imputed
2004 freshened spring-term seniors	—	Tested[2]
2002 sophomores: 2004 dropouts	Tested[1]	—
2002 sophomores: 2004 early graduates	Tested[1]	—
2002 sophomores: homeschooled in 2004	Tested[1]	—

— Not available.

[1] Imputed for base-year nonrespondents.

[2] Imputed for first follow-up participant test noncompleters.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

[18] An example of the latter is the link to the NCES Common Core of Data and Private School Survey provided via the NCES identification code (BYNCESSI). An analyst with a restricted-use license can import into the analysis such variables as, for example, *grade span* (highest grade and lowest grade of school for any of the relevant academic years); *percent minority*; proportion *free lunch qualifiers*; *enrollment*; grade 9 enrollment (2000–01), grade 10 enrollment (2001–02), grade 11 enrollment (2002–03), grade 12 enrollment (2003–04); *metropolitan status* (urbanicity): locale code; *student/teacher ratio*; *FTEs:* total number of full-time classroom teachers; *student enrollment:* overall; *school type* (regular, vocational, special education, other); and so on. A further example of such a restricted-use link is to school ZIP code, which permits locale variables to be imported from the 2000 decennial Census, and residential geocoding at the level of state, county, tract, and block. For the second follow-up, the link to the NCES Integrated Postsecondary Education Data System is especially important, and additional links to extant data have been supplied in the second follow-up, and are fully described in appendix N.

2.3.1 Test Design and Development

Test specifications for the ELS:2002 base year and first follow-up were adapted from frameworks used for NELS:88. There were two levels to the framework: content areas and cognitive processes. Mathematics tests contained items in arithmetic, algebra, geometry/ measurement, data/probability, and advanced topics (including analytic geometry and precalculus but not calculus). The tests also reflected cognitive process categories of skill/ knowledge, understanding/comprehension, and problem solving. The test questions were selected from previous assessments: NELS:88, NAEP, and PISA. Most of the base-year items were multiple choice (about 10 percent of the base-year mathematics items were open-ended). In the first follow-up, all items were multiple choice.

Both 10th-grade and 12th-grade items were field tested in 2001, and 12th-grade items were field tested again in 2003.[19] Items were selected or modified based on field test results. Final forms were assembled based on psychometric characteristics and coverage of framework categories.

The ELS:2002 assessments were designed to maximize the accuracy of measurement that could be achieved in a limited amount of testing time while minimizing floor and ceiling effects by matching sets of test questions to initial estimates of students' achievement. In the base year, this was accomplished by means of a two-stage test. In 10th grade, all students received a short multiple-choice routing test, scored immediately by survey administrators, who then assigned each student to a low, middle, or high difficulty second-stage form, depending on the student's number of correct answers in the routing test. In the 12th-grade administration, students were assigned to an appropriate test form based on their performance in 10th grade. Cut points for the 12th-grade low, middle, and high forms were calculated by pooling information from the field tests for 10th and 12th grades in 2001, the 12th-grade field test in 2003, and the 10th-grade national sample. Item and ability parameters were estimated on a common scale. Growth trajectories for longitudinal participants in the 2001 and 2003 field tests were calculated, and the resulting regression parameters were applied to the 10th-grade national sample. Test forms were designed to match the projected achievement levels of the lowest and highest 25 percent, as well as the middle 50 percent of the base-year sample 2 years later. An additional test form with a broad range of item difficulty was assembled for administration to follow-up participants who were new to the sample or who had not received a mathematics score in 10th grade. Additions to and deletions from the base-year sample resulted in 23 percent, 42 percent, and 26 percent of the follow-up sample taking the low, middle, and high difficulty forms, respectively, with the remaining 10 percent taking the broad-band form. Each of the four test forms contained 32 multiple-choice items.

[19] For more details about the field tests, see Burns et al. (2003) and appendix J of the *Base-Year to First Follow-up Data File Documentation*, Ingels et al. (2005b).

2.3.2 Assessment Framework for Mathematics

In the four tables immediately below (tables 3–6), content and process information[20] is provided about the 73 unique items that comprise the base-year and 59 items that comprise the first follow-up mathematics assessments. Additional tables are presented later that break down assignments of items by content and process *by test form*, and thus show the impact of overlap (any given unique item may appear on one or more forms).[21] Table 4 and table 5 show the numbers and percentages of unique mathematics test items devoted to each content area for the base-year and the first follow-up test batteries. Table 6 and table 7 show the number and percentages of unique test items devoted to each cognitive process area.

Table 3. Number and percentage of unique mathematics items in ELS:2002 base year, by content area: 2002

Content area	Number of items	Percentage of items
Arithmetic	19	26.0
Algebra	17	23.3
Geometry/measurement	20	27.4
Data analysis, statistics/probability	9	12.3
Advanced topics[1]	8	11.0

[1] "Advanced topics" includes precalculus and analytic geometry.

NOTE: To provide overlap, some items appear on more than one test form. Detail may not sum to totals because of rounding.

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

Table 4. Number and percentage of unique mathematics items in ELS:2002 first follow-up, by content area: 2004

Content area	Number of items	Percentage of items
Arithmetic	15	25.4
Algebra	17	28.8
Geometry/measurement	17	28.8
Data analysis, statistics/probability	4	6.8
Advanced topics[1]	6	10.2

[1] Advanced topics includes precalculus and analytic geometry.

NOTE: To provide overlap, some items appear on more than one test form. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

[20] Content by process (cognitive behavior) matrices can be useful for giving some sense of how tests have been constructed but must be interpreted with caution. Robitaille et al. (1993) point out that such grids somewhat oversimplify the interrelatedness of elements in the scheme. Analysts should consider that knowledge and abilities or behavior in one area of mathematics are not unconnected to knowledge and skills in other areas. As the National Assessment Governing Board has remarked on its 2005 NAEP mathematics framework (NAGB 2004), its divisions "are not intended to separate mathematics into discrete elements. Rather, they are intended to provide a helpful classification scheme that describes the full spectrum of mathematical content assessed by NAEP. Classifying items into one primary content area is not always clear cut, but doing so brings us closer to the goal of ensuring that important mathematical concepts and skills are assessed in a balanced way."

[21] There was also overlap across waves, in that some items were used both in the base year and first follow-up.

Table 5. Number and percentage of unique mathematics items per skill/cognitive process area in ELS:2002 base year, by process/skill specifications: 2002

Process/skill specifications	Number of items	Percentage of items
Procedural skills/knowledge	23	31.5
Conceptual understanding	27	37.0
Problem solving	23	31.5

NOTE: To provide overlap, some items appear on more than one test form. Detail may not sum to totals because of rounding.

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

Table 6. Number and percentage of unique mathematics items per skill/cognitive process area in ELS:2002 first follow-up, by process/skill specifications: 2004

Process/skill specifications	Number of items	Percentage of items
Procedural skills/knowledge	17	28.8
Conceptual understanding	26	44.1
Problem solving	16	27.1

NOTE: To provide overlap, some items appear on more than one test form. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

Table 7 shows the number of mathematics test items per form in the base year and first follow-up. Again, forms were assigned on the basis of performance on a routing test in the base year, but were assigned on the basis of the base-year ability estimate in the first follow-up. As earlier noted, those who had not been tested in the base year were given a broad range form in 2004. While all examinees received a 32-item form in 2004, the number of items ranged from 40 to 42 in the base year, except for a handful of students who received the single-stage 23-item version of the base-year assessment.

Table 7. Number of items in each ELS:2002 base-year and first follow-up test for assessing achievement in mathematics, by form: 2004

Form	Base year (2002)	First follow-up (2004)
Routing test	15	†
Second stage tests		
Form X (low difficulty)	25	32
Form Y (middle difficulty)	27	32
Form Z (high difficulty)	27	32
Form V (single stage in 2002; broad range in 2004)	23	32

† Not applicable.

NOTE: Some items overlap and appear on more than one test form.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002" and "First Follow-up, 2004."

While the tables above show the content and process areas for the unique items that comprise the overall base year and first follow-up mathematics tests, students took different forms of each test, and a given item could be used on more than one form. To see the number or proportion of items in a given content or skill area that students at various levels of form assignment in fact took, an additional set of tables is required. Table 8 shows content by

cognitive process distributions of items across all test forms. Contents of the routing tests are shown separately, although for purposes of computation of the base-year ability estimate, *theta*, the two stages of the mathematics test were combined.

Table 8. Number of mathematics items per content area, by cognitive skill/process and form, ELS:2002 base year through first follow-up: 2004

	Content area				
Cognitive skill/process	Arithmetic	Algebra	Geometry/ measurement	Data analysis/ statistics probability	Advanced topics[1]
Skill/knowledge					
Routing test	3	†	1	†	†
10th-grade low (X)	7	3	1	3	†
10th-grade medium (Y)	1	1	2	3	1
10th-grade high (Z)	†	2	1	†	†
10th-grade 1-stage (V)	2	3	1	†	†
12th-grade low (X)	7	4	2	†	†
12th-grade medium (Y)	2	4	1	†	1
12th-grade high (Z)	†	2	2	†	1
12th-grade broad (V)	4	3	2	†	1
Understanding/comprehension					
Routing test	1	4	1	1	†
10th-grade low (X)	3	†	1	1	2
10th-grade medium (Y)	2	3	2	1	5
10th-grade high (Z)	3	2	1	5	5
10th-grade 1-stage (V)	2	3	1	1	3
12th-grade low (X)	5	4	2	2	†
12th-grade medium (Y)	2	7	4	1	2
12th-grade high (Z)	†	5	4	1	4
12th-grade broad (V)	3	3	3	1	2
Problem solving					
Routing test	†	2	2	†	†
10th-grade low (X)	2	†	1	1	†
10th-grade medium (Y)	1	†	3	1	1
10th-grade high (Z)	1	1	10	1	†
10th-grade 1-stage (V)	2	†	3	1	1
12th-grade low (X)	2	†	3	1	†
12th-grade medium (Y)	2	1	5	†	†
12th-grade high (Z)	1	2	9	1	†
12th-grade broad (V)	3	2	4	1	†

† Not applicable.

[1] "Advanced topics" includes precalculus and analytic geometry.

NOTE: Some of the 73 base-year and 59 first follow-up items appear on more than one test form. The modal grade for sample members in 2004 was 12th grade; all sample members were 10th-graders in 2002.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002" and "First Follow-up, 2004."

Table 9 shows, by test form, numbers and percentage of items in each content area. The items in the base-year stage 1 test (routing test) have been combined with the items in the stage 2 test. For example, in the first follow-up (2004), students assigned the low form had 44 percent arithmetic items and no advanced topics, while students assigned the high form had 3 percent arithmetic items and 16 percent advanced topics. Nonetheless, the different forms comprise a

single test, and with IRT[22] methods, proficiencies can be estimated for ELS:2002 items not assigned to the examinee. In other words, all ELS:2002 IRT scores (whether number-right or probability of proficiency scores) measure student performance on the entire item pool regardless of which form they took.

Table 9. Percentage distribution of ELS:2002 test items, by content area and mathematics test form: 2002 and 2004

	Content area									
	Arithmetic		Algebra		Geometry/ measurement		Data analysis/ statistics/ probability		Advanced topics	
Mathematics test form	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number
10th-grade low form (X)	40.0	16	25.0	10	15.0	6	15.0	6	5.0	2
10th-grade medium (Y)	19.0	8	26.2	11	23.8	10	14.3	6	16.7	7
10th-grade high (Z)	11.9	5	31.0	13	38.1	16	7.1	3	11.9	5
10th-grade 1-stage (V)	26.1	6	26.1	6	21.7	5	8.7	2	17.4	4
12th-grade low (X)	43.8	14	21.9	7	25.0	8	9.4	3	0.0	0
12th-grade medium (Y)	18.8	6	37.5	12	31.3	10	3.1	1	9.4	3
12th-grade high (Z)	3.1	1	28.1	9	46.9	15	6.3	2	15.6	5
12th-grade broad (V)	31.3	10	25.0	8	28.1	9	6.3	2	9.4	3

NOTE: "Advanced topics" includes precalculus and analytic geometry. Detail may not sum due to rounding. Tenth-grade item summaries by forms X, Y, and Z combine the routing test and the second stage test. Twelfth grade was the model grade for sample members in 2004; all sample members were 10th-graders in 2002.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002" and "First Follow-up, 2004."

2.3.3 Assessment Framework for Reading

Reading items were drawn from two sources, NELS:88 and PISA (2000). There are four content areas:

- biographical;
- literary (including both poetry and prose);
- scientific (includes graphical displays as well as prose); and
- social studies.

There are three cognitive process areas: reproduction of detail, comprehension of thought (translating verbal statements into concepts), and inference/evaluative judgment (drawing conclusions based on the material presented). In the reading assessment (conducted in the base year only), there are 51 unique items, 11 of which are used twice (i.e., across two forms). Distribution of unique items (again, some items were repeated, to link forms) across the content areas is summarized in Table 10, while distribution across cognitive process areas is summarized in table 11.

[22] IRT stands for Item Response Theory. In ELS:2002, IRT was used both for vertical equating (linking the tests across grades 10 and 12) and lateral (or horizontal) equating (linking to HS&B in 1980 and to NELS:88 in 1990 and 1992). More generally, IRT is a test analysis procedure that applies mathematical models to the probability that any given examinee will provide a correct test response. Specifically, IRT uses patterns of correct, incorrect, and omitted answers to obtain ability estimates that are comparable across different test forms within a domain. In estimating a student's ability, IRT also accounts for each test question's difficulty, discriminating ability, and a guessing factor. For introductory information on IRT, see Embretson and Reise (2000) or Hambleton, Swaminathan, and Rogers (1991). For more technical discussions see Van der Linden and Hambleton (1997).

Table 10. Number and percentage of unique reading items in ELS:2002 base year, by content area: 2002

Content area	Number of items	Percentage of items
Biographical	12	23.5
Literary	18	35.3
Scientific	13	25.5
Social studies/other	8	15.7

NOTE: To provide overlap, some items appear on more than one test form. Detail may not sum to totals because of rounding.

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

Table 11. Number and percentage of unique reading items per skill/cognitive process area in ELS:2002 base year, by process/skill specifications: 2002

Process/skill specifications	Number of items	Percentage of items
Reproduction of detail	12	23.5
Comprehension of thought	19	37.3
Inferences/evaluative judgments	20	39.2

NOTE: To provide overlap, some items appear on more than one test form. Detail may not sum to totals because of rounding.

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

Again, the base-year reading test was a two-stage test in which a routing test guided examinees to the appropriate second-stage form. The number of items per first or second stage form is indicated in table 12.

Table 12. Number of items in each ELS:2002 base-year test form for assessing achievement in reading, by test form: 2002

Form	Number of items
Routing test	14
Second stage tests	
Form X (low difficulty)	16
Form Y (middle difficulty)	17
Form Z (high difficulty)	15

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

While the tables above show the content and process areas for the unique items that comprise the base-year reading assessment, students took different forms of the test, and a given item could be used on more than one form. To see the number or proportion of items in a given content or skill area that students at various levels of form assignment in fact took, an additional set of tables is required. Table 13 shows content by cognitive process distributions of reading items across all test forms. Contents of the routing tests are shown separately, although for computing the base-year ability estimate, *theta*, the two stages of the reading test were combined.

Table 13. Number of reading items per content area, by cognitive skill/process and form, ELS:2002 base year: 2002

Cognitive skill/process	Content area			
	Biographical	Literary	Scientific	Social studies
Reproduction of detail				
Routing test	†	1	2	†
10th-grade low (X)	†	3	†	2
10th-grade medium (y)	†	†	2	3
10th-grade high (Z)	1	†	1	1
Comprehension of thought				
Routing test	†	4	1	†
10th-grade low (X)	1	†	†	2
10th-grade medium (Y)	†	†	3	4
10th-grade high (Z)	6	†	1	2
Inferences and/or evaluative judgments				
Routing test	†	5	1	†
10th-grade low (X)	3	5	†	†
10th-grade medium (Y)	†	†	4	1
10th-grade high (Z)	1	†	1	1

† Not applicable.

NOTE: Some items appear on more than one test form.

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

Table 14 shows, by test form, numbers and percentage of items in each of the four reading content areas. The items in the base-year stage 1 test (routing test) have been combined with the items in the stage 2 test to show the total items examinees at each of the three levels were assigned.

Table 14. Percentage distribution of ELS:2002 test items, by content area and reading test form: 2002

Reading test form	Content area							
	Biographical		Literary		Scientific		Social studies	
	Percent	Number	Percent	Number	Percent	Number	Percent	Number
10th-grade low form (X)	13.3	4	60.0	18	13.3	4	13.3	4
10th-grade medium (Y)	†	†	32.3	10	41.9	13	25.8	8
10th-grade high (Z)	27.6	8	34.5	10	24.1	7	13.8	4

† Not applicable.

NOTE: Detail may not sum to totals due to rounding. Tenth-grade item summaries by forms X, Y, and Z combine the routing test and the second stage test.

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

2.3.4 Score Descriptions and Summary Statistics

Norm-referenced and criterion-referenced ELS:2002 test scores are explained below. For examples of the use of the ELS:2002 IRT-estimated number-right and probability of proficiency scores in cross-cohort analysis, see Cahalan et al. (2006). For an example of their use in

longitudinal analysis, see Bozick and Ingels (2007). For an example (from NELS:88) of use of a NAEP-scaled score, see Scott and Ingels (2007).

2.3.4.1 Norm-referenced Scores: Standardized Scores (T-scores)

The standardized scores (*theta* or T-scores) are overall measures of status at a point in time, but they are norm-referenced rather than criterion-referenced. They do not answer the question, "What skills do students have?" but rather, "How do they compare with their peers?" The transformation to a familiar metric with a mean of 50 and standard deviation of 10 facilitates comparisons in standard deviation units. For example, an individual with a T-score of 65 (or a subgroup with a mean of 65) has demonstrated achievement one and one-half standard deviations above the national average for 12th-graders, whereas a score of 45 would correspond to half a standard deviation below the norm. These numbers do not indicate whether students have mastered a particular body of material, but rather what their standing is relative to others. Base-year and first follow-up T-scores are documented in table 15.

Table 15. Standardized scores (*theta* or T-scores) from ELS:2002 mathematics and reading assessments, by variable: 2002 and 2004

Variable	Description	Range
BYTXMSTD	Base-year mathematics standardized score (T-score)	10–90
BYTXRSTD	Base-year reading standardized score (T-score)	10–90
BYTXCSTD	Composite mathematics + reading standardized score (T-score)	10–90
F1TXMSTD	First follow-up mathematics standardized score (T-score)	10–90

NOTE: T-score is the standardized score.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002" and "First Follow-up, 2004."

2.3.4.2 Norm-referenced Scores: Quartile Scores

Quartile scores divide the weighted (population estimate) achievement distributions into four equal groups, based on mathematics, reading, and mathematics plus reading composite scores. Quarter 1 corresponds to the lowest achieving quarter of the population, quarter 4 to the highest. Table 16 contains variable names, descriptions, and ranges for the quartile scores.

Table 16. Quartile scores from ELS:2002 mathematics and reading assessments, by variable: 2002 and 2004

Variable	Description	Range
BYTXMQU	Base-year mathematics quarter	1–4
BYTXRQU	Base-year reading quarter	1–4
BYTXCQU	Base-year composite mathematics + reading quarter	1–4
F1TXMQU	First follow-up mathematics quarter	1–4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002" and "First Follow-up, 2004."

2.3.4.3 Criterion-referenced Scores: IRT-estimated Number-right

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

type, or behavioral measures, and may be used in multivariate models as well. These scores may also be used as longitudinal measures of overall growth, when an aggregated measure is preferred. (When a disaggregated measure is desired, in order to measure and compare gains made at different points on the score scale [that is, to target a hierarchy of specific sets of skills], the probability of proficiency scores may be preferred in longitudinal analysis.)

For mathematics, 10th- and 12th-grade IRT-estimated number-right scores are available on both the ELS:2002 and the 1992 NELS:88 scale. Tenth-grade math scores are also available on the 1990 NELS:88 scale, to which 1980 HS&B scores can also be linked. For base-year reading, the scores are available on the NELS:88 scale as well as the ELS:2002 scale. The 1990 NELS:88 scale is documented in Ingels et al. (1994a,b) while the 1992 scale is documented in Rock and Pollack (1995). Linkage between NELS:88 and ELS:2002 was achieved through common item (anchor) equating. Tables 17 through 20 present IRT estimated number-right scores by variable, scale, and analysis. (See appendix D for errata regarding the first follow-up version of table 17.)

Table 17. ELS:2002 Item Response Theory (IRT)-estimated number-right reading and mathematics scores on the NELS:88 scale, by variable: 2002 and 2004

Variable	Description	Range	Weighted mean	Weighted standard deviation
BYNELS2R	Reading—NELS-equated estimated 10th-grade number-right (1992 scale)	0–54	29.2	9.5
BYNELS2M	Mathematics—NELS-equated estimated 10th-grade number-right (1992 scale)	0–81	44.4	13.7
BYNELS0M	Mathematics—NELS-equated estimated 10th-grade number-right (1990 scale)	0–58	37.6	11.4
F1NELS2M	Mathematics—NELS-equated 12th-grade estimated number-right (1992 scale)	0–81	50.1	14.2

NOTE: NELS:88 = National Education Longitudinal Study of 1988.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), “Base Year, 2002” and “First Follow-up, 2004.”

Table 18. Item Response Theory (IRT)-estimated number-right reading and mathematics scores on the ELS:2002 scale, by variable: 2002 and 2004

Variable	Description	Range	Weighted mean	Weighted standard deviation
BYTXRIRR	Reading IRT-estimated number-right	0–51	29.4	9.9
F1TXMBIR	Mathematics IRT-estimated number-right, 10th-grade, re-estimated on longitudinal scale	0–85	42.2	14.0
F1TXM1IR	Mathematics IRT-estimated number-right, longitudinal scale, all first follow-up participants	0–85	48.3	15.1
F1TXM1IR	Mathematics IRT-estimated number-right, longitudinal scale, first follow-up participants who were in 12th grade	0–85	48.6	15.1

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), “Base Year, 2002” and “First Follow-up, 2004.”

Table 19. Mathematics Item Response Theory (IRT)-estimated number-right scores, by analysis: 2004

Analysis	Scale	Variable
10th-grade cross-cohort (1980, 1990, 2002)	0–58	BYNELS0M
10th-grade cross-sectional (2002)	0–73	BYTXMIRR
10th-grade cross-cohort (1990–2002) (NELS scale)	0–81	BYNELS2M
10th-grade longitudinal NELS scale (2002–2004)[1]	0–81	BYNELS2M
10th-grade longitudinal ELS scale (2002–2004)[2]	0–85	F1TXMBIR
12th-grade longitudinal NELS scale (2002–2004)[1]	0–81	F1NELS2M
12th-grade longitudinal ELS scale (2002–2004)[2]	0–85	F1TXM1IR
12th-grade cross-cohort (NELS scale) (1992–2004)	0–81	F1NELS2M
12th-grade cross-sectional (ELS scale) (2004)	0–85	F1TXM1IR

[1] Use this pair in conjunction for gain measurement.

[2] Use this pair in conjunction for gain measurement.

NOTE: NELS = National Education Longitudinal Study of 1988. ELS=Education Longitudinal Study of 2002.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88); Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

Table 20. Reading Item Response Theory (IRT)-estimated number-right scores, by analysis: 2002

Analysis	Scale	Variable
10th-grade cross-sectional (2002)	0–51	BYTXRIRR
10th-grade cross-cohort (1990–2002) (NELS scale)	0–54	BYNELS2R

NOTE: NELS = National Education Longitudinal Study of 1988.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88); Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002."

2.3.4.4 Criterion-referenced Scores: Probability of Proficiency

Gains made at different points on the score scale have qualitatively different interpretations. For example, students who made 5-point gains by mastering arithmetical operations are learning very different lessons from those gaining 5 points at the high end of the scale by learning more advanced mathematics. Although the gains in number of scale score points may be the same, the interpretation, and the relationship with other factors such as coursework, can be expected to be quite different. For this reason, a continuous score representing the probability of proficiency at each of five mastery levels in mathematics and three mastery levels in reading was generated.[23]

Criterion-referenced proficiency probability scores are based on clusters of items that mark different levels on the reading and mathematics scales developed in NELS:88. Clusters of four items each were identified in the NELS:88 tests that marked three hierarchical levels in reading and five in mathematics. While clusters of four items anchor each mastery level, the probability of proficiency is a continuous score that does not depend on a student answering the actual items in each of the clusters but, rather, on the probability of a correct answer on these items given the overall pattern of response on the items completed. The three mastery levels for reading, and five for mathematics, are indicated below:

[23] For an illustration of the use of probability proficiencies in ELS:2002 math gain analysis, see Bozick and Ingels (2007). For further discussion of the nonequivalence of scale score points and consequent need (if achievement gain is to be fully interpreted) for multiple criterion-referenced proficiency levels that mark distinct learning milestones, see Rock (2007).

Probability of Proficiency, Reading Mastery Levels:

1. Simple reading comprehension, including reproduction of detail, and/or the author's main thought, such as identifying the objective of a character's action.
2. Simple inferences beyond the author's main thought and/or understanding and evaluating abstract concepts, such as identifying the author's state of mind, or inferring the meaning of a metaphor from context.
3. Complex inferences or evaluative judgments requiring multiple sources of information.

Probability of Proficiency, Mathematics Mastery Levels:

1. Simple arithmetical operations on whole numbers, such as simple arithmetic expressions involving multiplication or division of integers.
2. Simple operations with decimals, fractions, powers, and roots, such as comparing expressions, given information about exponents.
3. Simple problem solving, requiring the understanding of low-level mathematical concepts, such as simplifying an algebraic expression or comparing the length of line segments illustrated in a diagram.
4. Understanding of intermediate-level mathematical concepts and/or multistep solutions to word problems such as drawing an inference based on an algebraic expression or inequality.
5. Complex multistep word problems and/or advanced mathematics material such as a two-step problem requiring evaluation of functions.

The mastery levels are hierarchical in the sense that mastery of a higher level typically implies mastery at lower levels. The proficiency probabilities were computed using IRT-estimated item parameters calibrated in NELS:88. Each proficiency probability represents the likelihood that a student would pass a given mastery level defined as above in the NELS:88 sample. It should be remembered that probability of proficiency scores are IRT-derived estimates based on overall performance rather than counts of actual item responses. The NELS:88 and ELS:2002 tests were semi-adaptive, with different forms keyed to different ability levels. Owing to the multiple test forms used in NELS:88 and ELS:2002, not all sophomores received all items. Nevertheless, the IRT model permits proficiency probabilities to be estimated, even for those sophomores who were not administered a particular proficiency/mastery cluster. The mean of a proficiency probability score aggregated over a subgroup of students is analogous to an estimate of the percentage of students in the subgroup who have displayed mastery of the particular skill. Because the range of the scores is 0 to 1, means can be expressed in percentage form.[24] For example, the weighted mean for mastery of math level 1 in ELS:2002 is 0.92, which is equivalent to saying that 92 percent of the sophomore cohort had achieved mastery at this level (simple arithmetical operations on whole numbers). The probability of proficiency scores are summarized in table 21 (base year) and table 22 (first follow-up) below.

[24] On the interpretation of a probability as a proportion, see, for example, Fleiss, Levin, and Paik (2003, p. 1).

Table 21. Reading and mathematics probability of NELS-equated proficiency scores, by variable: 2002

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

NOTE: NELS = National Education Longitudinal Study of 1988.

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

Table 22. ELS:2002 Item Response Theory (IRT) NELS-equated mathematics proficiency probability scores: 2004

Variable name	Description	Range	Weighted mean	Weighted standard deviation
F1TX1MPP	Mathematics—level 1	0–1	0.96	0.12
F1TX2MPP	Mathematics—level 2	0–1	0.78	0.37
F1TX3MPP	Mathematics—level 3	0–1	0.62	0.45
F1TX4MPP	Mathematics—level 4	0–1	0.35	0.41
F1TX5MPP	Mathematics—level 5	0–1	0.04	0.14

NOTE: NELS = National Education Longitudinal Study of 1988.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

2.3.4.5 Psychometric Properties of the Tests

Information about the psychometric properties of the test items, the setting of difficulty levels, differential item functioning, and scoring procedures, are provided in the two field test documents (Burns et al. 2003, chapter 5, and Ingels et al. 2005b, appendix J). IRT scaling and linking procedures follow the NELS:88 precedent, using a 3-parameter IRT model in PARSCALE (Muraki and Bock 1991); the NELS:88 procedure is described in Rock and Pollack (1995).

Reliabilities were computed using the variance of the posterior distribution of plausible values for each test-taker's *theta* (ability estimate), compared with the variance of the *thetas* across the whole sample (i.e., error variance versus total variance). The reliability estimates are the proportion of "true variance" (1 minus error variance) divided by total variance (see Samejima [1994] on this procedure).

For the combined base-year and first follow-up tests, the reliability was 0.92. This reliability is a function of the variance of repeated estimates of the IRT ability parameter (within-variance), compared with the variability of the sample as a whole (Ingels et al. 2005b). This 0.92 reliability applies to all scores derived from the IRT estimation. Imputed test scores were not included in the reliability estimation.

The use of IRT-scale scores and the adaptive testing approach used in ELS:2002 limit the concern that gain scores may be unreliable due to floor and ceiling effects.

2.3.4.6 Indicators of Student Motivation at Both Testing Points

One major concern in measuring achievement is whether students are motivated to do their best on low-stakes tests, such as the mathematics assessment in ELS:2002. This concern may be particularly strongly felt with reference to spring-term seniors, who may be in the process of disengaging from high school in anticipation of the transition to postsecondary education or the work force, and who may have had their fill of assessments, in the form of such high-stakes tests as exit exams and college entrance exams. Although the greatest concern may be felt about spring-term seniors, concerns about motivation rightly encompass high school sophomores as well.

While there is no single definitive measure of student motivation on the tests, there are several possible indicators of the comprehensiveness and quality of the test data collected. For example, in scoring the 2002 and 2004 tests, the assessment subcontractor examined "pattern marking"[25] and missing responses. They did not find evidence of pattern marking, nor high levels of omitted items. For example, in the ELS:2002 first follow-up with around 11,000 mathematics assessments completed, 17 assessments were discarded for these reasons: 11 test records were deleted because tests were incomplete (fewer than 10 items answered) and 6 more because response patterns indicated lack of motivation to answer questions to the best of the student's ability. In the base year, 10 assessments were discarded for incompleteness, and none for pattern marking.

Given that participation in the survey was voluntary, and that a student could have opted to not participate, or to participate by completing the questionnaire only, the student response rate may also be an indirect indicator of student test-taking motivation. Generally NAEP sees a drop in participation in grade 12, compared to grades 4 and 8. For ELS:2002's predecessor study, NELS:88, lower participation rates were registered in 12th grade as well.[26]

For the ELS:2002 base year, the weighted participation rate was 87 percent. Of the 15,362 participants, 95 percent (weighted) also completed the test. (Some who did not complete the test could not be validly tested for language or disability reasons.)

For the ELS:2002 first follow-up (2004), when most sample members were high school seniors, the overall participation rate increased slightly from the base year, to a weighted 89 percent. Of the test-eligible questionnaire completers, some 87 percent (weighted) of

[25] An example of "pattern marking" would be responses of "A" for all answers or ABCABCABC through most or all of the test. Patterned responses such as "11111111…" or "12345432123454321…" or "1515151515…" can be identified by a simple algorithm sequentially comparing the difference between each test item and the next one, and calculating the variance of the absolute differences. In the first example given, the inter-item differences are always zero, in the second, always 1 or -1, and in the third, 4 or -4. In each case, the variance of the absolute differences is equal to zero, whereas for four- or five-choice test items, the variance of absolute differences for motivated respondents tends to be close to 1.0. All tests with variances of less than .5 were reviewed and those few with identifiable pattern marking were deleted.

[26] Fully interpreting the senior year decline in test completion in NELS:88 is difficult. There was sample dispersion, and the policy was to test transfer students, though the resources for doing so were limited. In consequence, often a questionnaire might be completed over the telephone and the test sacrificed, despite the student's willingness to be assessed. In contrast, in ELS:2002, transfers were ineligible for the first follow-up test and did not count against the assessment response rate—however, test scores were imputed for all transfers. No test score imputation was undertaken in NELS:88. Because studies such as NELS:88 and ELS:2002 induct their initial samples prior to 12th grade, they may be less affected by a "senioritis" phenomenon, in that students have already committed to the study and may have developed a sense of membership in the panel. Certainly for HS&B, the prior longitudinal cohort study that in its sophomore cohort most closely resembles ELS:2002 in design, participation was higher in the modally 12th-grade first follow-up than in the 10th-grade base year (and higher than the 12th-grade participation rate for the HS&B senior cohort that was selected in the same schools in 1980).

questionnaire completers also completed the test. Looking specifically at questionnaire completion for senior cohort members who remained in the same school at both points in time, a 97 percent survey participation rate was obtained both overall and for each race/ethnicity groups (Ingels et al. 2005b, table 39). If voluntary participation rates are to some degree indicative of student motivation, then there is some evidence that seniors may have taken the assessment seriously.[27] The overall pattern—lack of high numbers of omitted response, lack of "pattern-marking," high test reliability,[28] and high participation rates in both rounds of the study—argue for the credibility and quality of the test data. In short, while lack of motivation for some students surely affected test results in ways that could not be identified and edited out, most test takers answered all or almost all the items, and internal-consistency reliabilities were high for all subgroups examined, both in the field tests and full-scale studies. These are good indications that interpretation of test results in the aggregate should not be significantly compromised by low test-taking motivation.

2.3.4.7 Score Linkages With External Assessments: NAEP and PISA

The ELS:2002—NAEP 12th-grade linkage. One new assessment variable has been produced subsequent to the release of the first follow-up student data in 2005. More specifically, the ELS:2002 12th-grade mathematics test has been linked to 12th-grade NAEP. The 2004 ELS:2002 first follow-up mathematics tests did not share common items with the NAEP 2005 mathematics assessment. As a result, common item equating was not possible, so score scales were linked by means of an equipercentile transformation.

Equating—"the process of developing a conversion from the system of units of one form of a test to the system of units of another form so that scores derived from the two forms after conversion will be equivalent and interchangeable" (Angoff 1982)—is the strongest form of test linkage. It ensures that the scores that are linked are truly equivalent and statistically and conceptually interchangeable. However, a variety of stringent conditions must be met to successfully equate. These conditions include essential alikeness in content such that the two tests are congeneric (i.e., they measure the same underlying factor); the tests must measure the same populations; they should be of similar reliability; they should meet the condition of equity (it should be a matter of indifference to the result which test examinees take); and they should be symmetric (the function equating X to Y should be the inverse of the function equating Y to X) (see Kolen and Brennan 2004; Linn 1993; Lord 1980; Mislevy 1992).

Arguably, NAEP and ELS:2002 mathematics content is quite similar, and both tests attempt to measure the same underlying factor. At grade 12 in 2005, NAEP's primary emphasis was on geometry/measurement and algebra (NAGB 2004). This is also the case for the ELS:2002 tests in both the high and medium form. (Of course the ELS:2002 assessment is individually adaptive, and for this reason, understandably, there is proportionately more arithmetic [number properties and operations, in NAEP parlance] and less geometry in the ELS:2002 low form [taken by the bottom 23 percent of examinees] than in the NAEP test.)

[27] Note that ELS:2002 sample members were given a cash incentive for participation. The effects of payment on test-taking motivation are unknown. Because test reliabilities were high and incomplete tests and pattern marking did not seem to be a problem, one interpretation might be that students made a reasonable effort, regardless of whether they did so out of a sense of obligation for being paid to do a task or for more idealistic reasons.

[28] Imputed test scores were not included in the calculation of reliabilities.

The tested populations are also highly similar—spring-term high school seniors—though not identical (ELS:2004 tested 2004 seniors, and NAEP 2005 seniors). However, there are also many important differences between ELS:2002 and NAEP that impact the linking procedures and interpretation of linked scores. Though test content is similar, item formats were somewhat different (a mixture of free response and multiple choice for NAEP, but only the latter for the ELS:2002 12th-grade math test). While both the NAEP and ELS:2002 mathematics assessments are highly reliable, they achieve this end through different means (the ELS:2002 tests assigned different forms to candidates of different ability; NAEP, on the other hand, includes auxiliary information in calculating the posterior estimates of ability). Nonetheless, the NAEP design is driven by the need to maximize reliability for group-level measurement, and, unlike ELS:2002 scores, NAEP scores are not designed to be reliable at the individual level (Beaton and Gonzalez 1995). The condition of equity (that examinees should be indifferent as to which test they take) is difficult to meet given the difference between an adaptive test in ELS:2002 and a test based on a matrix sample of items in NAEP. Finally, scoring methods differed in several respects, and may particularly have affected the ability to transform the scores in the tails of the distribution.

The NAEP-ELS:2002 linked mathematics score should therefore be described as a concordance[29] rather than an equating. Though the scores may be comparable (there is a linkage that is based on distributional similarities), no claim is made that the scores may be treated as equivalent (that is, that they have precisely the same meaning). The NAEP-scale score represents the score level achieved by students of the same percentile rank in two populations that were matched as closely as was possible given the differences in sample (e.g., only ELS:2002 12th-graders were used in the linking exercise). Linking scales to yield concordant scores relies on minimal assumptions about the comparability of the tests involved (on concordance, see Dorans 2004 and Pommerich and Dorans 2004). Neither means, standard deviations, reliabilities, nor standard errors of measurement are assumed to be the same. The tests need only be roughly congeneric in that they measure essentially the same basic underlying factor.

Linking procedures for the ELS:2002 NAEP-scaled math score. To maximize the likeness of the two linking samples, a subsample of ELS:2002 students was used to compute equivalent percentiles. Transformations were computed based only on the subset of ELS:2002 first follow-up participants who were in 12th grade in spring 2004 (using the "G12COHRT" flag to select cases, and the "F1QWT" weight to generalize to the national population of 12th-graders).

The equipercentile transformation was carried out using 3-moment smoothing of the weighted frequency distributions. Plots of the equipercentile-equated scores showed extreme deviations in the tails of the distribution from a trend line based on linear approximation. These deviations are probably due to the methodology employed in NAEP scoring: the NAEP scores are transformations of normally distributed IRT ability estimates, which if no shrinkage is imposed, tend to have long tails. The ELS:2002 scores, which are sums of probabilities, do not. As a result, the equipercentile conversion becomes distorted in the tails of the distributions. Throughout most of the score range, a 1-point difference in ELS:2002 mathematics scale corresponds to a difference of about 2.25 points in the NAEP metric. But in the extreme tails of the distribution a 1-point difference in ELS:2002 mathematics score corresponds to a difference of up to 4 points in the NAEP metric. Although these distortions occur only for a small number of students, a combination of the equipercentile transformation and a linear approximation of the

[29] Another test score concordance appears on the ELS:2002 second follow-up data files, a concordance between sample members' ACT and SAT scores.

transformation was used to assign scores. The cut points for using equipercentile versus linear transformation were selected such that the ELS:2002 to NAEP link would be monotonic, and are indicated in table 23.

Table 23. Linking methods for implementing NAEP high school senior mathematics scales in ELS:2002/2004, by scale score range: 2004/2005

ELS scale score range	Linking method	Weighted percent of data
15.20–27.49	Linear approximation	10.5
27.50–79.39	Equipercentile transformation	89.1
79.40–82.54	Linear approximation	0.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004" and National Assessment of Educational Progress (NAEP), 2005.

The result of the linking exercise is the variable F1TXNAEP, a NAEP-scaled version of the ELS:2002 IRT-estimated number right score (F1TXM1IR). While the historical NAEP vertical scale has been expressed in a 0–500 range, NAEP 2005 12th-grade mathematics results have not been vertically scaled with 8th- and 4th-grade results, and are on a 12th-grade scale of 0–300. (The ELS:2002 scale has a range of 0–85.)

As further documentation of the linkage, sample differences in weighted population estimates were reviewed for each survey. Percentages of racial/ethnic groups were quite similar (given slightly different definitions). Detecting whether the small differences are due to sampling variability or adjustments, or other factors such as differences in race/ethnicity classification schemes, is impossible (table 24).

Table 24. Comparison of ELS:2002 and NAEP 2005 12th-grade mathematics linking samples, by sex and race/ethnicity: 2004/2005

ELS:2002 (2004)—Grade 12			NAEP (2005)—Grade 12		
Sex and race/ethnicity	Population	Weighted percent	Sex and race/ethnicity	Population	Weighted percent
Total	2,996,374	100.0	Total	2,877,208	100.0
Sex			Sex		
Male	1,494,597	49.9	Male	1,382,104	48.0
Female	1,501,777	50.1	Female	1,495,103	52.0
Race/ethnicity[1]			Race/ethnicity[1]		
American Indian or Alaska Native	28,375	1.0	American Indian or Alaska Native	27,709	1.0
Asian or Pacific Islander	134,933	04.5	Asian or Pacific Islander	146,698	5.1
Black or African American	399,745	13.3	Black or African American	390,286	13.6
Hispanic or Latino	450,727	15.1	Hispanic or Latino	385,519	13.4
More than one race	117,420	3.9	More than one race	—	—
Unclassified	—	—	Unclassified	21,193	0.7
White and all other races	1,865,174	62.3	White and all other races	1,905,802	66.2

— Not available.

[1] "White and all other races" is predominantly White, with a very small number of individuals from other race categories. All race categories exclude individuals of Hispanic or Latino origin.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004" and National Assessment of Educational Progress (NAEP), 2005.

Note that while the equating sample was restricted to ELS:2002 spring-term 2004 12th-graders, once the transformation of ELS:2002 to the NAEP scale was determined, NAEP-scaled

scores could also be assigned for ELS:2002 first follow-up participants who were not high school seniors (for example, 2002 sophomores who were held back a grade between 2002 and 2004), making the NAEP-scaled score available for all 13,702 sample members with an ELS:2002 first follow-up mathematics score.

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

PISA linkage. In addition to the NAEP linkage, two further external linkages were carried out, both of them with PISA—reading (2000) and mathematics (2003). The PISA-scaled reading score is BYPISARE; the PISA-scaled math score is BYPISAME. For full documentation of these linkages, see Ingels et al. (2004, 2005b).

2.4 High School Transcript Component; Course Offerings File

The ELS:2002 high school transcript data collection sought information about coursetaking from students' official high school records (e.g., courses taken while attending secondary school, credits earned, year and term a specific course was taken, and final grades). When available, other information, such as dates enrolled, reason for leaving school, and standardized test scores such as ACT and SAT[30] results, was collected. Because of the size and complexity of the file and the reporting variation by school, additional variables were constructed from the raw transcript file to facilitate analyses. These variables include standardized grade point averages, academic "pipeline" measures, and total credits earned by subject area. The construction of many of the transcript variables is based on Carnegie units. A Carnegie unit is equal to a course taken every day, one period per day, for a full school year. All transcript items and composite variables have been appended to the ELS:2002 restricted-use data files and require special access for individual analysis. However, summary variables, such as Carnegie units in the main academic subjects, have been included on the ECB and the Data Analysis System (DAS).

In addition to high school transcripts, information is also provided about the course offerings of the base-year schools. For analysis purposes, school course offering information can be attached to the student record.

2.5 Second Follow-up Questionnaire Content

A single web-based instrument was developed for ELS:2002 second follow-up sample members, in which the respondents could self-administer the interview or complete it assisted by a telephone interviewer or field interviewer. In all modes of administration, *the identical web-based instrument* was accessed. This approach eliminated the potential for mode of

[30] Transcript-reported SAT and ACT scores have been augmented in the second follow-up by additional scores obtained through records-matching with the test developers. Data from the multiple sources were merged, and an SAT-ACT concordance was created, so that both sets of scores would be on a common scale.

administration effects due to differences in question wording or response options.[31] Also, content areas most susceptible to interviewer (versus self-administration) effects, such as sensitive items with high potential for eliciting social desirability biases, were largely avoided. Finally, the instrument design process took into account the need to ensure that items would reflect similar levels of cognitive demand across modes (e.g., formats requiring extensive visual information to be easily understood would not be appropriate, since visual cues could not be provided in a telephone interview).

The instrument development process was launched with a meeting of the study's TRP in August 2005. Panelists recommended that the full-scale interview capitalize on the study's rare opportunity to examine the transition from high school to postsecondary education. Project instrument development staff were urged to concentrate on issues related to college access and choice in this round of the study. The project team reworked the field test instrument, consulting with experts in postsecondary education as needed. Instrument items were drawn from a number of studies including Baccalaureate and Beyond, Beginning Postsecondary Student Longitudinal Study, HS&B, NELS:88, and the National Postsecondary Student Aid Study.

The interview was organized into four substantive sections: *High School*, *Postsecondary Education*, *Employment,* and *Community*. The interview concluded with a *Locating* section. Appendix E includes flowcharts for each of the four substantive sections of the interview. They document the sequence of questions and the web-based instrument's routing logic. A facsimile of the instrument, also found in appendix E, documents question wording and response options. An in-depth description of each of these sections follows.

The first section, *High School*, collected retrospective information about high school completion. The majority of respondents skipped this section entirely because their high school completion date and the type of high school credential they earned were preloaded into the instrument at the start of data collection. The preloaded information was drawn from high school transcripts when available or from the first follow-up early graduate (see F1S15 and F1E27) and dropout (see F1D41 and F1D45) interviews. The high school transcript data were still undergoing quality control procedures when the second follow-up data collection began. In an effort to preload only stable transcript data, transcript information was only preloaded for cases where the following conditions were met: (1) the high school completion date was May or June 2004, the modal dates of completion; (2) the credential was a high school diploma or a certificate of attendance; and (3) quality control had been completed.[32] In summary, second follow-up respondents were asked whether they had completed high school, the date they had completed high school, and the credential earned if they had not already provided this information in a first follow-up interview and any one of the following conditions were met: (1) their high school transcript was not collected, (2) their high school transcript data (at the start of data collection) indicated that they had completed high school in a month other than May or June 2004, (3) their high school transcript data (at the start of data collection) indicated they had earned a GED, or

[31] Of course, eliminating these two sources of mode effects is not to say that mode effects could not have occurred (for example, on the basis of differences such as self- versus interviewer administration). However, methodological work with similar items, age groups, and populations in the NCES postsecondary longitudinal studies (which also employ both web self-administration and computer-assisted interviewer administration) has not uncovered mode effect problems (see, for example, the following NCES methodology and field test reports: NCES 2004-02, NCES 2006-01, and NCES 2005-02).

[32] Despite this effort, the preloaded transcript information was later determined to be incorrect for some of the cases. Consequently, the preloaded data do not match the final released transcript data for a small number of cases. F2PHSDG indicates the credential earned as it was preloaded. The preloaded high school completion dates are found in F2PHSDT.

(4) their high school transcript data (at the start of data collection) indicated that a high school credential had not been awarded by the high school(s) providing transcripts. As will be discussed in greater detail where appropriate, high school completion dates, as preloaded or reported in the interview, played an important role in instrument routing logic and composite variable construction (see section 7.2.2.1).[33]

A second important purpose of the *High School* section was to retrospectively classify respondents as spring-term 2004 12th-graders, spring-term 2004 dropouts, neither, or for a small set both (see G12COHRT and F2SP04DO). The spring term of 2004 is of interest as this was the reference period for the first follow-up data collection. For a more detailed description of the classification procedures see section 7.2.2.1.

First follow-up nonrespondents who were identified as spring-term 2004 dropouts as well as those identified as early alternative completers (earned a GED prior to April 2004) were asked a series of retrospective questions about why they had dropped out of high school prior to or during the spring term of 2004. These questions were repeated from the first follow-up dropout and early graduate interviews. Responses to these items from the first follow-up and the second follow-up interviews are combined in composite variables (see F2WYLV1–F2WYLV14).

First follow-up questions about the GED were also repeated in the second follow-up *High School* section. All second follow-up respondents who reported earning a GED since they were last interviewed were asked a series of questions on the topic of their high school credential. Like the questions related to dropping out of high school, data collected from both rounds of the study were combined in composite variables (see F2GEDPRG, F2GEDOTH, F2GEDST, and F2WYGED1–F2WYGED6).

Questions in the *High School* section of the interview also identified a small set of respondents who were attending high school in the spring term of 2006 (F2RTYPE = 6). Many of the questions in the remainder of the interview, particularly those related to postsecondary education, did not pertain to these individuals. Therefore, these high school students were not asked to answer the majority of the questions in the *Postsecondary Education* section and select questions thereafter.

The *Postsecondary Education* section of the interview, the point of entry for most respondents, focused on education *after* high school. Questions pertained to the application process, admissions, financial aid offers, institutions attended, experiences at these institutions, and educational expectations. Retrospective information about dual enrollment experiences at postsecondary institutions during high school was not collected.

Since the primary focus of this interview is the transition out of high school, respondents who submitted applications more than once, as for example, to transfer from one postsecondary institution to another, were asked to identify only those postsecondary institutions they had applied to as part of their *first* round of applications. For the same reason, the first postsecondary institution the respondent attended after high school received special attention in a series of questions (see F2PS1, F2B13A–F, F2B14, F2B15, F2B16A–C, F2B17A–D, F2B18A–G). In most cases, the school of interest in these questions was the postsecondary institution with the

[33] The data user is cautioned that many of the variables that provide data as it was collected in the *High School* section of the second follow-up interview, that is, variables with an "F2A" prefix, are not standalone variables to be used in analyses. They serve as inputs to composite variables only. They are only provided on the ECB for reference or validation of composite variable construction.

earliest enrollment date after high school completion or exit. In cases for which enrollment in a fall-term postsecondary institution was immediately preceded by summer school attendance, the fall-term institution was selected as the first.[34]

Complete month-by-month enrollment histories for all postsecondary institutions attended after high school were collected in the *Postsecondary Education* section. These enrollment histories in conjunction with the date of high school completion or exit, as preloaded or reported in the *High School* section of the interview, were used to classify respondents into one of six mutually exclusive categories (see F2RTYPE): Standard enrollees, Delayers, Leavers, Delayer-Leavers, Nonenrollees, and High School students. Table 25 indicates the characteristics of each respondent type.

Table 25. Classification rules for F2RTYPE, by respondent type: 2006

Respondent type	Any postsecondary enrollment after high school?	"On time" postsecondary enrollment?	Any reported postsecondary enrollment in 2006?	Enrolled in high school when interviewed?
Standard enrollee	Yes	Yes	Yes	No
Delayer	Yes	No	Yes	No
Leaver	Yes	Yes	No	No
Delayer-leaver	Yes	No	No	No
Nonenrollee	No	†	†	No
High school student	†	†	†	Yes

† Not applicable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Broadly speaking, respondents may be divided into those who have attended a postsecondary institution after high school and those who have not. First we will address those who reported some postsecondary enrollment following high school. *Standard enrollees* were respondents who enrolled in a postsecondary institution "on time," that is, within the first enrollment window following their high school completion or exit date[35] and had some postsecondary enrollment in 2006 prior to the date of their interview. *Delayers* were enrollees who started their postsecondary education *after* the first enrollment window following their high school completion or exit date[36] and had some postsecondary enrollment in 2006 prior to the date of their interview. *Leavers* were enrollees who began their postsecondary education "on time," but had no postsecondary enrollment in 2006 prior to the date of their interview. Note that leavers did not necessarily drop out of their postsecondary program. Leavers may have completed a postsecondary credential. *Delayer-leavers* were both delayers and leavers.

[34] These questions pertained to a fall-term postsecondary institution following summer school enrollment when the following conditions were met: (1) the respondent completed high school, enrolled in a summer school (in May, June, or July), ended summer school (in May, June, July, or August), and enrolled in a postsecondary institution for the fall term (in August, September, or October) within the same calendar year; and (2) the earliest and most recent dates of enrollment at the fall-term postsecondary institution spanned a greater number of months than the dates of enrollment at the summer school.

[35] Respondents who completed or dropped out of high school from January through July were considered "on time" if they began their postsecondary education by October of the same calendar year. Respondents who completed or dropped out of high school from August through December were considered "on time" if they began their postsecondary education by the following February.

[36] Respondents who completed or dropped out of high school from January through July were classified as delayers if they did not begin their postsecondary education by October of the same calendar year. Respondents who completed or dropped out of high school from August through December were classified as delayers if they did not begin their postsecondary education by the following February.

Respondents in the remaining two categories had no postsecondary enrollment following high school. The vast majority had completed or dropped out of high school. These respondents were classified as *Nonenrollees*. As noted previously, a small number of respondents reported that they were still enrolled in high school. These respondents are identified as *High schoolers*.

Table 26 illustrates which questions associated with various postsecondary education topics were administered to each respondent type. All respondents, with the exception of high school students, were asked if they had applied to a postsecondary institution since high school. Those who reported that they had were asked follow-up questions about those applications, whether those applications were accepted, and the financial aid offers received. All of these post-high school respondents, regardless of whether they reported applying to a postsecondary institution, were asked whether they had attended a postsecondary institution following high school. Respondents who indicated they had not were then classified as nonenrollees. All others were then asked to name the institution(s) they had attended and provide the dates of their enrollment. Based on these enrollment dates and the date of their high school completion or exit, enrollees were subdivided into the standard enrollees, delayers, leavers, and delayer-leavers as described previously. The remaining postsecondary education topic areas and the respondent types to which they relevant are listed in table 26.

Table 26. Administration of postsecondary education topics, by respondent type: 2006

Respondent type	Standard enrollee	Delayer	Leaver	Delayer-leaver	Nonenrollee	High school student
Whether has applied	Yes	Yes	Yes	Yes	Yes	No
Questions about applications	If applicable	If applicable	If applicable	If applicable	If applicable	No
Whether was accepted	If applicable	If applicable	If applicable	If applicable	If applicable	No
Questions about offers	If applicable	If applicable	If applicable	If applicable	If applicable	No
Whether has attended	Yes	Yes	Yes	Yes	Yes	No
Enrollment history	Yes	Yes	Yes	Yes	No	No
Reasons for delaying	No	Yes	No	Yes	No	No
Reasons no longer enrolled	No	No	Yes	Yes	No	No
Why took a break from postsecondary enrollment	If applicable	If applicable	If applicable	If applicable	No	No
Why attended part-time	If applicable	If applicable	If applicable	If applicable	No	No
Why switched postsecondary institutions	If applicable	If applicable	If applicable	If applicable	No	No
Questions about first postsecondary institution	Yes	Yes	Yes	Yes	No	No
Major at 2006 postsecondary institution	Yes	Yes	No	No	No	No
Financing post-secondary education	Yes	Yes	Yes	Yes	No	No
Reason has not attended a postsecondary institution	No	No	No	No	Yes	No
Educational expectations	Yes	Yes	Yes	Yes	Yes	Yes

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

The administration of each of the five topic areas in the *Employment* section was also determined by the respondent type classification scheme (see table 27). The initial two topics pertained to employment nonconcurrent with postsecondary education. The questions in the first of these two sets of questions referred to the first job after high school. Delayers, delayer-leavers, and nonenrollees were eligible for these items since these respondents all had a significant period of time after high school when they were not enrolled at a postsecondary institution. The second employment module focused on employment at the time of the interview. Nonenrollees, delayer-leavers, leavers, and high schoolers were subject to this module because they were not enrolled at a postsecondary institution at the time of the interview. The next set of questions focused on jobs held by postsecondary students during the 2004–05 and 2005–06 academic years. All four types of postsecondary enrollees were eligible for these questions if their postsecondary attendance coincided with these academic years. In contrast, only nonenrollees were eligible for the next topic. They were questioned about months of unemployment when a gap existed between high school and their first job, their first job and their current job, and/or their first job and the date of the interview if they were not currently working. Based on these responses as well as the employment dates provided in the first two modules, month-by-month employment status variables were constructed beginning with June 2004 (see F2EM0206-F2EM0608–

F2EM0608). Most of the remaining questions in the *Employment* section pertained to all respondent types. Topics included income, finances, and occupational expectations at age 30.

Table 27. Administration of employment topics, by respondent type: 2006

Respondent type	Standard enrollee	Delayer	Leaver	Delayer-leaver	Nonenrollee	High school student
First job	No	Yes	No	Yes	Yes	No
Current job	No	No	Yes	Yes	Yes	Yes
Unemployment history	No	No	No	No	Yes	No
Postsecondary student jobs	If applicable	If applicable	If applicable	If applicable	No	No
Finances/occupational expectations	Yes	Yes	Yes	Yes	Yes	Yes

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

The final substantive section of the interview, *Community*, covered topics related to family formation, living arrangements, community involvement including military service, and experiences that may influence the life course. With one minor exception, all questions pertained to all respondent types. The interview concluded with the *Locating* section which collected information that will be used to contact the respondents in the next round of the study. Since these data are not provided on the ECB or the DAS, documentation for this section is not provided.

A complete list of variables provided on the second follow-up ECB and DAS is presented in appendix L. See appendix N for a discussion of ancillary data that were collected in the second follow-up to augment sample members' records. Several sources of extant data were tapped including the College Board, ACT, ACE GED testing service, and federal loan and grant databases. Appendix M lists second follow-up composite variables with brief descriptions. A descriptive overview of composite variables constructed from second follow-up data and extant data is provided in chapter 7 (section 7.2.2).

Chapter 3
Sample Design

3.1 Base-Year and First Follow-up Sample Design

3.1.1 Overview

This chapter describes the Education Longitudinal Study of 2002 (ELS:2002) base-year, first follow-up, and second follow-up sample designs, including the design of the first follow-up high school transcript component.

Section 3.1 provides a historical summary of sample design issues for the base year and first follow-up. Starting with section 3.2, the chapter provides an expanded discussion of the sample design in the context of the ELS:2002 second follow-up in 2006.

The ELS:2002 base-year sample design comprises two primary target populations—schools with 10th grades and sophomores in those schools—in the spring term of the 2001–02 school year. ELS:2002 used a two-stage sample selection process. First, schools were selected. These schools were then asked to provide sophomore enrollment lists, from which students were selected.

Schools and students are the study's basic units of analysis. School-level data reflect a school administrator questionnaire, a library media center questionnaire, a facilities checklist, and the aggregation of student data to the school level. Student-level data consist of student questionnaire and assessment data and reports from students' teachers and parents. (School-level data, however, can also be reported at the student level and serve as contextual data for students.)

The basis for the sampling frame for the first follow-up (2004) was the sample of schools and students used in the ELS:2002 base-year sample. There are two slightly different target populations for the first follow-up. One population consists of those students who were enrolled in the 10th grade in 2002. The other population consists of those students who were enrolled in the 12th grade in 2004. The former population includes students who dropped out of school between 10th and 12th grades, and such students are a major analytical subgroup, as are transfer students. Note that in the first follow-up, a student is defined as a member of the student sample; that is, an ELS:2002 spring 2002 sophomore or a freshened first follow-up spring 2004 12th-grader. In the first follow-up, high school transcripts were also collected. The basis for the transcript sample was all student sample members who had participated in either the 2002 base year, the 2004 first follow-up, or both.

3.1.2 Base-Year Sample Design

The sample design for ELS:2002 is similar to the designs used in the three prior studies of the National Education Longitudinal Studies Program: the National Longitudinal Study of the High School Class of 1972 (NLS:72), the High School and Beyond (HS&B) longitudinal study, and the National Education Longitudinal Study of 1988 (NELS:88). ELS:2002 is different from NELS:88 (but similar to HS&B) in that the ELS:2002 base-year sample students are 10th-graders rather than 8th-graders. As in NELS:88, Hispanics and Asians were oversampled in

ELS:2002. However, for ELS:2002, counts of Hispanics and Asians were obtained from the Common Core of Data and the Private School Survey to set the initial oversampling rates.

ELS:2002 used a two-stage sample selection process. First, schools were selected with probability proportional to size.[37] School contacting resulted in 1,221 eligible public, Catholic, and other private schools from a population of approximately 27,000 schools containing sophomores. Of the eligible schools, 752 agreed to participate in the study. These schools were then asked to provide sophomore enrollment lists. In the second stage of sample selection, approximately 26 students per school were selected from these lists. Additional information on the base-year sample design can be found in chapter 3 and appendix J of the base-year data file user's manual (Ingels et al. 2004, NCES 2004-405).

The target population of schools for the ELS:2002 base year consisted of regular public schools, including state Department of Education schools and charter schools, and Catholic and other private schools that contained 10th grades and were in the United States (the 50 states and the District of Columbia).

The sampling frame of schools was constructed with the intent to match the target population. However, selected schools were determined to be ineligible if they did not meet the definition of the target population. Responding schools were those schools that had a Survey Day (i.e., data collection occurred for students in the school).[38] Of the 1,268 sampled schools, there were 1,221 eligible schools and 752 responding schools (68 percent weighted participation rate).

A subset of most but not all responding schools also completed a school administrator questionnaire and a library or media center questionnaire (99 percent and 96 percent weighted response rates, respectively). Most nonresponding schools or their districts provided some basic information about school characteristics, so that the differences between responding and nonresponding schools could be better understood, analyzed, and adjusted. Additionally, RTI field staff completed a facilities checklist for each responding school.

The target population of students for ELS:2002 consisted of spring-term sophomores in 2002 (excluding foreign exchange students) enrolled in schools in the school target population. The sampling frames of students within schools were constructed with the intent to match the target population. However, selected students were determined to be ineligible if they did not meet the definition of the target population. Of the 19,218 sampled students, there were 17,591 eligible sophomores. The 15,362 participants on the public-use file represent a weighted student response rate of 87 percent.

The ELS:2002 base-year survey instruments comprised two assessments (reading and mathematics) and a student questionnaire. Participation in ELS:2002 was defined by questionnaire completion. Although most students were asked to complete the assessment battery in addition to the questionnaire, there were some cases in which a student completed the questionnaire but did not complete the assessments. Guidelines were provided to schools to assist them in determining whether students would be able to complete the ELS:2002 survey instruments.

[37] The size used was a composite measure of size based on school enrollment by race/ethnicity. See Appendix J of the *Base Year Data File User's Manual* (Ingels et al. 2004) for more details.
[38] One eligible school had no eligible students selected in the sample. This school was considered a responding school.

Students who could not complete the ELS:2002 questionnaire (by virtue of limited English proficiency or physical or mental disability) were part of the expanded sample of 2002 sophomores who were followed in the study and eligibility status was reassessed 2 years later. There were 163 such students. To obtain additional information about their home background and school experiences, contextual data were collected from the base-year parent, teacher, and school administrator surveys.

The student sample was selected, when possible, in the fall or early winter so that sample teachers could be identified and materials could be prepared well in advance of Survey Day. However, selecting the sample in advance meant that some students transferred into the sample schools and others left between the time of sample selection and Survey Day. To address this issue, sample updating was conducted closer to the time of data collection. Complete enrollment lists were collected at both the time of initial sampling and the time of the sample update.

One parent of the sample student and English and mathematics teachers of the sample student were also included in the base-year sample.

3.1.3 First Follow-up Sample Design

There are two target populations for the ELS:2002 first follow-up. Because of these two target populations and the major analytical subgroups, the sample included the following types of students:

- ELS:2002 base-year student respondents who were currently enrolled in either the 12th grade or some other grade in the school in which they were originally sampled. All such students were included in the first follow-up sample.
- ELS:2002 base-year student respondents who finished high school early, including those who graduated from high school early, as well as those who obtained alternative certification (e.g., exam-certified equivalency such as the General Educational Development credential). All such students were included in the first follow-up sample.
- ELS:2002 base-year sample students who were deemed unable to participate during the base year owing to disability or insufficient command of the English language. All such students were included in the follow-up sample.
- ELS:2002 base-year student respondents who dropped out of school prior to data collection in the 12th grade. All such students were included in the follow-up sample.
- ELS:2002 base-year student respondents who transferred out of the school in which they were originally sampled, including those who transferred to a homeschool setting. All such students were included in the follow-up sample.
- A subsample was included of base-year nonrespondents (including those who did not have parental consent). Some base-year nonrespondents had remained at the base-year school, while others finished high school early, transferred, or were dropouts or homeschooled in spring term 2004.
- Students at the base-year sample school who were enrolled in the 12th grade but who were not in 10th grade in the United States during the 2002 school year. During spring term 2002 such students may have been out of the country, enrolled in school

in the United States in a grade other than 10th, had an extended illness or injury, been institutionalized, been homeschooled, or temporarily dropped out of school. A spring term-based "freshening" sample of such students was included in the first follow-up.

If a base-year school split into two or more schools and ELS:2002 base-year sample members moved *en masse* to a new school, the study followed them to the destination school and sought the school's participation in the first follow-up. These schools can be thought of as additional base-year schools in a new form. Specifically, a necessary condition of adding a new school in the first follow-up was that it arose from a situation such as the splitting of an original base-year school, thus resulting in a large transfer of base-year sample members (usually to one school, but potentially to more). Four base-year schools split, and five new schools were spawned from these four schools. At these new schools, as well as at the original base-year schools, students were tested and interviewed. Additionally, the 12th-grade sample was freshened, and the administrator questionnaire administered.

3.1.3.1 Eligibility

All spring-term 2002 sophomores in eligible schools (i.e., schools that matched the target population as defined in section 3.1.1), except for foreign exchange students, were eligible for the base-year study. Base-year-eligible students were assumed to again be eligible in the first follow-up, regardless of school enrollment status. Additionally, all spring-term 2004 seniors in the base-year schools, except for foreign exchange students, were eligible for the first follow-up. Some base-year students were out of scope for the first follow-up (but sometimes were in-scope again in the second follow-up). Reasons for being temporarily (for the particular round of data collection) out of scope included being institutionalized or out of the country and thus unavailable through the data collection period. Reasons for being permanently out of scope included mortality and correction of sampling errors in which a noncohort member had been mistakenly selected.

Several categories of students who were ineligible for HS&B and NELS:88[39] were eligible for ELS:2002 (though it did not mean that such students were necessarily tested or that they completed questionnaires). In NELS:88, the following categories of students were deemed ineligible:

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

In ELS:2002, such students were deemed (test and) *questionnaire-incapable*, while remaining eligible for the sample. Base year contextual data were collected for such students (who appear only on the restricted-use files), and their eligibility status was reassessed in the first follow-up. Some students could be administered a questionnaire but could not complete a test. Students

[39] For a summary of ineligibility and exclusion issues in HS&B and NELS:88 see Ingels (1996).

deemed capable of responding to a questionnaire but not capable of completing an assessment[40] were treated as regular sample members.

3.1.3.2 First Follow-up Subsampling

A base-year nonrespondent student was defined as a student who was selected in the base year and did not complete a student questionnaire. For the first follow-up, a subsample of 1,000 nonrespondent students was selected from the 2,229 base-year nonrespondents. Initially, a subsample of 1,620 nonrespondents was selected. All nonresponding students were included with certainty (i.e., probability equal to one), except for White students in public schools who were randomly subsampled. Then, to help the response rate and to conserve resources, the subsample of 1,620 was randomly subsampled across all student types to 1,000 nonrespondents.

3.1.3.3 Sample Freshening

Because part of the target population consists of those students who were enrolled in the 12th grade in the spring of 2004, the first follow-up included students at the base-year sample school who were enrolled in the 12th grade in the spring[41] of 2004 but who were not in the 10th grade in the United States during the spring of 2002. During this time, such students may have been out of the country or may have been enrolled in school in the United States in a grade other than 10th (either at the sampled school or at some other school). In addition, some students may have reenrolled, although in spring 2002 they were temporarily out of school, owing to illness, injury, institutionalization, homeschooling, or school dropout. Some 238 new students were added to the study under the freshening procedure, although 31 of the 238 were incapable of completing the questionnaire.

The total sample for the public-use file in the first follow-up comprised 16,515 individuals of whom 14,989 participated for a weighted response rate of 88.7 percent.

3.1.3.4 High School Transcript Study Sample Design

In autumn 2004, high school transcripts were requested for all sample members who participated in at least one of the first two student interviews: the base-year interview or the first follow-up interview. Thus, sample members who were dropouts, freshened sample members, transfer students, homeschooled students, and early graduates are included if they were respondents in either the 2002 or 2004 interview. Transcripts were also requested for students who could not participate in either of the interviews because of a physical disability, a mental disability, or a language barrier. Further information about the transcript component may be found in Bozick et al. 2006 (NCES 2006-338), available to licensed users of the transcript data.

3.2 Second Follow-up Sample Design

The target populations of the ELS:2002 second follow-up (2006) were the 2002 sophomore cohort and the 2004 senior cohort. The sophomore cohort consists of those students

[40] For example, a student with vision problems might not be able to complete a written test, but might be able to respond to an interviewer's oral administration of a questionnaire.

[41] The ELS:2002 cohorts, like the NAEP 12th-grade samples and the prior high school longitudinal cohorts (NLS:72, HS&B, and NELS:88), are spring-defined. For ELS:2002 this means that fall-term 2003 12th-graders who were not in 10th grade in the United States 2 years before (spring term 2002) and were not in 12th grade in the spring term of 2004 are not represented in the sample. Such individuals would normally be either fall-term dropouts or fall-term 2003 early graduates.

who were enrolled in the 10th grade in the spring of 2002 and the 12th-grade cohort comprises those students who were enrolled in the 12th grade in the spring of 2004. The sophomore cohort includes students who were in the 10th grade in 2002 but not in the 12th grade in 2004 (i.e., sophomore cohort members but not senior cohort members). The senior cohort includes students who were 12th-graders in 2004 but were not in the 10th grade in U.S. schools in 2002; they were included through a sample freshening process as part of the first follow-up activities.

The basis for the ELS:2002 second follow-up sampling frame was the sample of students selected in the base year when they were 10th graders in 2002 combined with the sample of freshened students who were in the 12th grade in 2004.

Figure 2 shows the distribution of the approximately 17,600 eligible students sampled from 750 schools in the base year (BY) plus the 240 students added during freshening in the first follow-up.[42] For the first follow-up full-scale study, there were a total of 18,000 eligible sample members that included 15,400 BY respondents, 2,200 BY nonrespondents, 160 questionnaire-incapable[43] BY students, 210 freshened students, and 30 questionnaire-incapable freshened students.

For the second follow-up full-scale study, there were 17,900 eligible sample members who included all first follow-up eligible sample members except deceased students (approximately 20), study-ineligible[44] members (approximately 10), and base year nonrespondents or freshened sample members who were out-of-scope sample members in the first follow-up study (about 20). The second follow-up fielded sample consisted of 16,400 sample members (see figure 2) as follows:

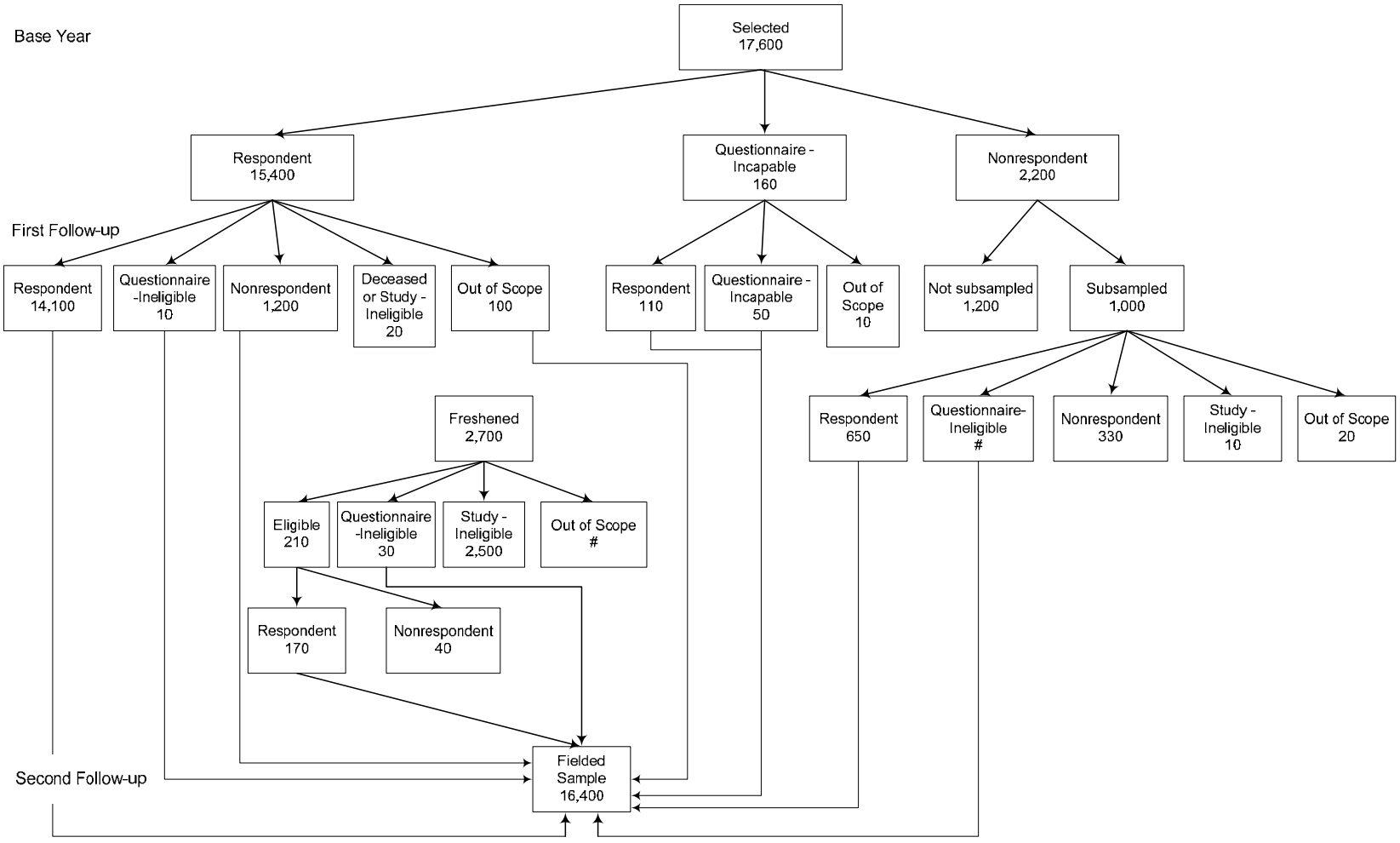
- respondents for both the BY and F1 rounds (14,100);
- F1 nonrespondents who were BY respondents (1,200);
- BY nonrespondents who were subsampled in the F1 and responded in the F1 (650);
- BY or F1 questionnaire-incapable members (210);
- freshened respondents in F1 study (170); and
- BY respondents who were determined to be out-of-scope in the F1 (100).

The sample members listed above made up the second follow-up sample that was fielded, but there were some prior-round nonrespondents who, while eligible members of one or both of the ELS:2002 target populations, were not fielded.[45] These nonrespondents included the following types of sample members:

- BY nonrespondents who were also nonrespondents in the F1 study;

[42] Readers are reminded that second follow-up sample sizes for subgroups are approximate. There was no public-use data file for the second follow-up. Exact sample sizes from restricted-use data files cannot be published unless the data are perturbed in some way. The perturbation approach taken here was to round the exact sample sizes of cells to tens or hundreds.
[43] Questionnaire-incapable students were ineligible for questionnaire or test completion owing to language barriers or severe disabilities but were included in the sample; contextual data were collected for them, and their eligibility status reassessed.
[44] Study-ineligible sample members are individuals who were not members of the relevant cohort (2002 sophomores or 2004 freshened seniors) but were initially included owing to sampling error and subsequently reclassified as permanently out of scope.
[45] The nonfielded sample members who were base year-first follow-up nonrespondents or first follow-up freshening sample nonrespondents were treated as eligible sample members classified as nonrespondents for the weighting adjustments and in the nonresponse bias analysis.

Figure 2. ELS:2002 second follow-up full-scale sample: 2006



Rounds to zero.

NOTE: "Study-Ineligible" means not a member of the spring-term 2002 sophomore cohort and not a member of the spring 2004 senior cohort for freshening; or, ineligible by virtue of being a foreign exchange student. All sample sizes have been rounded. Numbers of less than four digits have been rounded to tens. Numbers of four or five digits have been rounded to hundreds.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002," "First Follow-up, 2004," and "Second Follow-up, 2006."

- freshened nonrespondents; and
- sample members who asked to be removed from the study.

Some 330 base-year and first follow-up nonrespondents and 40 freshened nonrespondents were not fielded since lack of base-year and first follow-up high school information for these sample members meant that these sample members would have no analytical value in the full-scale study.[46] A handful of sample members who asked to be removed from the study were treated as permanent nonrespondents. The sample excluded members who were determined to be study-ineligible in either the base year or the first follow-up, such as sample members who are deceased (whose ineligibility begins with their date of death) or were sampled in error based on cohort membership information later found to be erroneous.

Once fielded, some members of the sample of 16,400 were determined to be out of scope. There were 460 out-of-scope second follow-up sample members, who fell into five basic groups, as indicated in table 28.

Table 28. Numbers of out-of-scope cases in the second follow-up, by out-of-scope reason: 2006

Out-of-scope reason	Number
Deceased	40
Out of country	210
Institutionalized/incarcerated	50
Questionnaire incapable/incapacitated	80
Unavailable for duration of 2006 data collection	80

NOTE: Numbers are rounded to tens.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Apart from the deceased, these individuals are regarded as temporarily out of scope only. If available for future interviews, they will be asked to participate. The portion of the sample that is out of scope is in flux across rounds. However, sample members are more likely to temporarily out of scope, as they disperse after high school and assume new roles, including roles in the military or work force that may take them out of the country or otherwise render them inaccessible. (Military personnel could fall into any of the above categories; their status cannot be separately distinguished.)

[46] There were also 1,200 base-year nonrespondents who were sampled out of the study prior to first follow-up data collection, with another 1,000 base-year nonrespondents retained. The subsample of 1,000 base-year nonrespondents became the basis for nonfielded sophomore cohort cases in the second follow-up, in those instances in which they were nonrespondents in the first follow-up as well.

Chapter 4
Data Collection Results and Methodology

4.1 Base-Year and First Follow-up Data Collection Results

This chapter briefly describes data collection for the Education Longitudinal Study of 2002 (ELS:2002) base-year and first follow-up surveys and, more expansively, data collection for the second follow-up. The discussion of the first follow-up includes data collection for the high school transcript component as well as information about the administration of the test and questionnaires.

More detailed accounts of the base-year and first follow-up data collections can be found in the following NCES publications:

- *Education Longitudinal Study of 2002: Base Year Data File User's Manual* (Ingels et al. 2004; NCES 2004-405);
- *Education Longitudinal Study of 2002: Base-Year to First Follow-up Data File Documentation* (Ingels et al. 2005b; NCES 2006-344); and
- *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation* (DFD)[47] (Bozick et al. 2006; NCES 2006-338).

Base-year data were collected in spring term 2002. The base-year survey collected data from students, parents, teachers, librarians, and school administrators. Pre-data-collection activities included securing endorsements from educational organizations and gaining cooperation from state education agencies, school districts, and individual schools. Self-administered questionnaires and achievement tests were the principal research instruments. Data collection primarily took place during in-school survey sessions conducted by an RTI field survey administrator.

First follow-up data were collected in spring term 2004, from students (including transfers) as well as dropouts; transcripts were collected in the next school year.

A total of 752 high schools participated in the base year, resulting in a weighted school response rate of 67.8 percent. School cooperation results are set out in table 29. Response and coverage rates for base-year and first follow-up student and student-contextual components (including transcript coverage) are provided in tables 30 through 33.

[47] The transcript DFD report (NCES 2006-338) is available only to licensed users of the transcript data; however, substantial attention is given to the transcript component in the present document as well.

Table 29. Unweighted school sampling and eligibility, and unweighted and weighted participation, by sampling stratum: 2002

School sampling stratum	Sampled schools		Eligible schools		Participating schools		
	Number	Unweighted percent[1]	Number	Unweighted percent[2]	Number	Unweighted percent[3]	Weighted percent
Total	1,268	100.0	1,221	96.3	752	61.6	67.8
School sector							
Public	953	75.2	926	97.2	580	62.6	69.1
Catholic	140	11.0	140	100.0	95	67.9	74.0
Other private	175	13.8	155	88.6	77	49.7	62.9
Urbanicity							
Urban	434	34.2	414	95.4	250	60.4	67.3
Suburban	630	49.7	609	96.7	361	59.3	59.8
Rural	204	16.1	198	97.1	141	71.2	79.3

[1] Percent is based on overall total within column. Details may not sum to 100 percent due to rounding.

[2] Percent is based on number sampled within row.

[3] Percent is based on number eligible within row.

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

A total of 15,362 students participated, primarily in in-school sessions, for an 87.3 percent weighted response rate.[48] In addition, each sampled student's mathematics teacher and English teacher were given a questionnaire to complete. Weighted student-level coverage rates for teacher data were 91.6 percent (indicating receipt of a report from either the math teacher, the English teacher, or both). School administrators and library media coordinators also completed a questionnaire (weighted response rates were 98.5 percent and 95.9 percent, respectively). Mail questionnaires were sent to parents with a telephone follow-up for nonresponders. Student coverage for parent questionnaires was 87.5 percent (weighted). RTI survey administrators completed a facilities checklist at each school. The number of completed instruments and both weighted and unweighted response rates are summarized in table 30.

[48] In a two stage-sample, a final response rate should be viewed as the product of both levels of participation. For example, with a school response rate of 67.8 percent and a student response rate of 87.3 percent, the final response rate taking both stages of the design into account is 67.8 * 87.3 = 59.2 percent. A school nonresponse analysis was conducted in the base year to establish that nonresponse bias at the school level was minimal and to provide a fuller basis for nonresponse adjustments in the final weighting. Similar analysis and adjustment were undertaken at the student level. For details see Ingels et al. (2004), *Education Longitudinal Study of 2002: Base Year Data File User's Manual* (NCES 2004-405), chapter 3, section 3.2.6.

Table 30. Summary of ELS:2002 base-year response and coverage rates, by instrument: 2002

Instrument	Selected	Participated	Weighted percent	Unweighted percent
Student questionnaire	17,591	15,362	87.3	87.3
Student assessment[1]	15,362	14,543	95.1	94.7
Parent questionnaire[2]	15,362	13,488	87.5	87.8
Teacher ratings of students[3]	15,362	14,081	91.6	91.7
School administrator questionnaire	752	743	98.5	98.8
Library media center questionnaire	752	718	95.9	95.5
Facilities checklist	752	752	100.0	100.0

[1] Percentage of cases for which a student questionnaire and cognitive test were obtained. When a test was not obtained, test results were imputed.

[2] Indicates a coverage rate: the proportion of participating students with a parent report. More parents participated; completed case numbers reflect the records in the public-use data file, where parent (and teacher) data were excluded for students who did not complete a base-year student questionnaire.

[3] Indicates a coverage rate: ratings obtained from at least one teacher.

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

First follow-up in-school data collection occurred between January and June 2004. Out-of-school data collection took place between February and August 2004 and included telephone and in-person interviews. Results are summarized in table 31.

Table 31. Summary of ELS:2002 first follow-up response and coverage rates, by instrument: 2004

Instrument	Selected	Participated	Weighted percent	Unweighted percent
Total sample for public-use file	16,515	14,989	88.7	90.8
Student questionnaire	13,092	12,427	93.4	94.9
Student assessment[1]	12,427	10,995	87.4	88.5
School administrator questionnaire[2]	12,427	11,856	95.9	95.4
Transfer questionnaire	1,799	1,275	68.4	70.9
Dropout questionnaire	876	686	73.2	78.3
Early graduate questionnaire	687	560	80.6	81.5
Homeschool questionnaire	61	41	61.5	67.2

[1] Indicates a coverage rate: percentage of cases for which a student questionnaire and cognitive test were obtained. When a test was not obtained, test results were imputed.

[2] Indicates a coverage rate: percentage of students affiliated with base-year (2002) schools in 2004 (student questionnaire completers) for whom a school administrator report was obtained.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

Student questionnaire completers comprise those participating first follow-up sample members then currently (spring term 2004) associated with a base-year (2002) school. In other words, the student questionnaire sample was drawn from base-year sophomore cohort members who remained at their base-year school or seniors brought in through the freshening process at those same schools. There were 13,092 individuals in the sample eligible to complete a student questionnaire, and 12,427 did so. The overall response rate for this group was 93.4 percent, weighted (94.9 percent unweighted).

The mathematics assessment was administered to about 87 percent (weighted) of the student questionnaire sample (again, the individuals who remained in, or were freshened in, the

base-year schools). For this same sample (students associated with a base-year school 2 years later), school administrator data are available 95 percent (weighted) of the time.

Not all sophomore cohort members remained in their base-year schools. Many transferred to a new school. These students completed a transfer student questionnaire. (Although they did not complete the mathematics assessment, a mathematics score was imputed for them.) For transfer students, a 68.4 percent weighted (70.9 percent unweighted) response rate was achieved.

Dropouts were defined in ELS:2002 as sample members who were absent from school for 4 consecutive weeks or more at the time of the survey, and not absent due to accident or illness. The weighted sophomore cohort dropout participation rate was about 73 percent (over 78 percent unweighted).

For all sample types (including questionnaire-incapable students), high school transcripts were also collected in the first follow-up, in the course of the 2004–05 school year. About 91 percent of sample members had a complete or incomplete[49] transcript. Table 32[50] provides information about transcript coverage overall and by selected subgroups. Table 33 breaks out coverage information by cohort as well as subgroup.

[49] Note that some transcript records were necessarily incomplete (for example, the transcripts of a dropout, or of a student who repeated a year between the two surveys), while other records may be incomplete (especially for transfers) because complete information could not be obtained.

[50] Note that because first follow-up transcript data (and second follow-up questionnaire data) are available in restricted-use electronic codebooks (ECBs) only (supplemented by a public-use Data Analysis System), sample size information has been perturbed, by a process of rounding, as an additional protection against inadvertent or deductive disclosure of respondents' identifying information. Because a public-use ECB was produced for the ELS:2002 base year and first follow-up (other than the transcript component), precise sample sizes for the public-use file (which differs slightly in number from the restricted use files [e.g., questionnaire-incapable sample members do not appear on the public-use files]) appear in text and tables describing the 2002 and 2004 rounds. Exact sample sizes are also provided for the second follow-up field test (2005); field test data are not released, even in restricted form, and therefore pose no danger of deductive disclosure.

Table 32. Percentage of base-year and first follow-up students with a complete or incomplete transcript, by selected characteristics: 2004–05

Student characteristic	Rounded sample size	Weighted percent	Unweighted percent
Total	16,400	90.7	91.1
Sex			
Male	8,200	89.9	90.9
Female	8,200	91.4	91.4
Race/ethnicity[1]			
American Indian or Alaska Native	140	92.4	90.8
Asian or Pacific Islander	1,700	90.7	90.8
Black or African American	2,200	88.3	87.5
Hispanic or Latino	2,500	86.9	89.6
More than one race	800	91.4	91.2
White and all other races	9,100	92.2	92.5
School sector			
Public	12,900	90.6	90.6
Catholic	2,000	95.0	94.8
Other private	1,500	86.0	90.6
Urbanicity			
Urban	5,500	86.8	88.6
Suburban	7,900	92.7	93.0
Rural	3,000	91.3	90.9
School region[2]			
Northeast	3,000	83.3	85.7
Midwest	4,100	91.8	92.6
South	6,000	91.2	91.0
West	3,400	94.3	94.3

[1] "White and all other races" is predominantly White, with a very small number of individuals from other race categories. All race categories exclude individuals of Hispanic or Latino origin.

[2] Region is defined by the U.S. Census Bureau based on the state in which the school is located.

NOTE: Detail may not sum to totals because of rounding. Because the transcript file is restricted use only, sample sizes have been rounded, and are thus approximate.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "High School Transcript Component."

Table 33. Percentage of base-year and first follow-up students with a complete or incomplete transcript, by grade cohort and selected characteristics (weighted): 2004–05

Student characteristic	Cross-sectional		Panel
	10th-grade (G10) cohort[1] (student weight, F1TRSCWT) (unweighted, N = 16,170)	12th-grade (G12) cohort[2] (student weight, F1TRSCWT) (unweighted, N = 13,420)	10th- to 12th-grade panel (student weight, F1PNLWT) (unweighted, N = 13,250)
Total	90.6	93.1	93.1
Sex			
Male	89.8	92.7	92.7
Female	91.4	93.5	93.5
Race/ethnicity[3]			
American Indian or Alaska Native	92.3	94.6	94.5
Asian or Pacific Islander	90.5	92.7	92.4
Black or African American	88.2	91.8	91.8
Hispanic or Latino	86.9	90.0	90.0
More than one race	91.4	94.2	94.2
White and all other races	92.1	94.1	94.0
School sector			
Public	90.6	93.1	93.1
Catholic	94.9	95.5	95.5
Other private	85.6	90.8	90.6
Urbanicity			
Urban	86.7	89.9	89.8
Suburban	92.7	94.8	94.7
Rural	91.3	93.5	93.5
School region[4]			
Northeast	91.8	86.6	94.0
Midwest	83.2	94.0	86.5
South	91.2	93.6	93.6
West	94.3	96.7	96.7

[1] G10 cohort indicates the cross-sectional population of the nation's 2002 spring-term sophomores.

[2] G12 cohort indicates the cross-sectional population of the nation's 2004 spring-term seniors.

[3] "White and all other races" is predominantly White, with a very small number of individuals from other race categories. All race categories exclude individuals of Hispanic or Latino origin.

[4] Region is defined by the U.S. Census Bureau based on the state in which the school is located.

NOTE: Because the transcript file is restricted use only, sample sizes have been rounded, and are thus approximate.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "High School Transcript Component."

4.2 Base-Year and First Follow-up Data Collection Methods

Although the results of base-year and first follow-up data collection have been described above, section 4.2 describes pre-data-collection and data collection activities—the basic data collection methodology followed in the in-high-school years of the study.

4.2.1 Base-Year Data Collection Methodology

Before public school recruitment could begin, it was necessary to obtain permission to contact the schools, first from the states and then from districts. For Catholic schools, permission was sought at the diocesan level, while other private schools were contacted directly, without intermediary.

Schools were initially contacted by mail, with a package of materials about the study. Several days after the package was sent, the school was contacted by telephone. If the school agreed to participate, a school coordinator was identified to serve as a point of contact and to help handle the logistical arrangements for the survey. Dates for a Survey Day and two Makeup Days were scheduled. At the same time, staff members were designated to receive the school administrator and library media center questionnaires. It was determined whether the type of parental consent used by the school was active (written) or passive (implicit). Schools were offered the opportunity to provide endorsement letters to be included with the consent letter to the parents. Every effort was made to "convert" noncooperating schools. Nonetheless, there were substantial numbers of refusals; indeed, about 38 percent (unweighted) of the contacted eligible schools refused to participate.

In each cooperating school, the coordinator was asked to provide an enrollment list of 10th-grade students, which was used as the basis for sample selection. Since some students may have transferred into or out of the school's 10th grade over subsequent weeks, the sample was updated before the Survey Day, with new students given a chance of selection into the sample.

The actual survey session was conducted by RTI staff, as a group administration, for all students who wished to participate and whose parents had given their implied or explicit consent. First, students were given a timed routing test in math and reading. After completing the routing tests, the students completed the student questionnaire. While the students completed the questionnaire, the survey administrators graded the routing tests and used the resulting scores to determine which of the second-stage test forms in math and reading (low, medium, high ability) to assign to each student. While the students completed the second-stage tests, RTI survey administrators edited the student questionnaires for completeness by checking critical items and attempting to retrieve missing information or clarify ambiguities.

The routing test was allotted 12 minutes in math and 14 minutes in reading. The second stage test was 18 minutes for math and 16 minutes for reading. The questionnaire was to be completed in 45 minutes.

If less than 100 percent of the eligible students participated on Survey Day, the RTI survey administrator attempted to confirm the Makeup Day that had been scheduled during the school recruitment process. Of the 15,362 participants, 85.4 percent were surveyed in their school on Survey Day, another 11.1 percent were surveyed on a Makeup Day, and 3.5 percent were surveyed outside school over the telephone.

School administrator and librarian questionnaires were also collected. Survey administrators completed a facilities checklist that evaluated the school's physical plant and safety features. Finally, by the end of the data collection period, at least one teacher report had been received for 92.4 percent of all of the participating students.

In addition to surveys of within-school populations, a parent survey was conducted. Parent questionnaires were mailed on or soon after the school's scheduled Survey Day to all

parents for whom addresses had been obtained through the school. For parents with no address available, the parent questionnaire was not mailed until the student questionnaire was sent in and the locator information (which included home address) was recorded. Parents returned the questionnaire to RTI in a postage-paid envelope. RTI staff followed up with nonresponding parents by telephone and in person. Of the 15,362 responding students, parent data (either by mailed questionnaire or by telephone interview) were received from 13,488 of their parents for a weighted coverage rate of 87.4 percent.

4.2.2 First Follow-up Test and Questionnaire Data Collection Methodology

States and districts had been informed in the base year that there would be another study round 2 years hence. For districts, a courtesy letter was mailed reminding them about ELS:2002 and stating that their schools would be contacted to gain permission to collect follow-up data.

Some 752 schools participated in the base-year study (although one had no eligible selected 10th-graders). When base-year schools were recontacted for the first follow-up, it was learned that five of the schools no longer had sample members (enrolled in any grade at the school) or high school seniors (hence no freshening sample). These schools, therefore, were no longer eligible for the study. Of the eligible schools, 698 (93.4 percent) allowed RTI to return to collect data in the schools. In 44 cases, the school refused to allow RTI to return to the school to collect data. Three districts (representing a total of five schools) also refused to allow RTI to return to their schools to collect data. Data from students enrolled at these schools/districts were collected outside of the school setting. Students at the base-year schools completed student questionnaires and a math test at the in-school administration. School administrator questionnaires were collected. A handful of base-year schools split into multiple schools between 2002 and 2004. Thus, in addition to schools that participated in the base year, five schools that received pools of students from base-year schools were included as new schools in survey activities but were not added to the probability sample.

In the spring and again in the autumn of 2003, each base-year school was provided a list of ELS:2002 base-year sample members from its school. The school was asked to indicate whether each sample member was still enrolled at the school. For any sample member who was no longer enrolled, the school was asked to indicate the reason and date the student left. If the student had transferred to another school, the base-year school was asked to indicate the name and location of the transfer school. This information was gathered again in the spring of 2004, prior to the school's scheduled Survey Day. In the fall of 2003, each base-year school was also asked to provide a list of the 12th-graders enrolled at that school, so this information could be used as part of the freshening process.

As earlier noted, schools were asked to identify sample members who no longer attended the base-year school. At the time, contact information for those individuals was collected. However, further tracing of sample members was often required, using (when available) the locating information provided by parents and students in the base year.

The in-school survey sessions were essentially similar to those in the base year. However, there was no reading test. In addition, while there were multiple test forms each tailored to ability level, the math test form was not assigned on the basis of a routing test as in the base year. In the first follow-up, the math test form was assigned on the basis of the prior (base-year) test score, as was done in the National Education Longitudinal Study of 1988 (NELS:88). For the mathematics

assessment, 26 minutes was allotted, with 45 minutes for the student questionnaire. A school administrator questionnaire was also administered in the first follow-up, and course offerings information was collected for base-year schools as well. Of those who participated in in-school survey sessions, 87.5 percent (9,737) were interviewed on Survey Day, and the remaining 12.5 percent on a Makeup Day. Some 10.1 percent (1,126) participated on the first Makeup Day, and 2.4 percent (262) on a subsequent Makeup Day. Of course, for schools that did not allow a Makeup Day, students were pursued outside of the school setting.

Not all spring 2002 sophomores remained in their base-year schools. Some had dropped out of high school; others had transferred. A few shifted to a homeschool setting, while others graduated early. Therefore, a large segment of the data collection took place outside the school setting. No attempt was made to test students who had transferred out of their base-year schools by 2004; however, test scores were imputed for this group. For students not in their original schools, telephone data collection began in February 2004. For sample members under the age of 18, parental permission was obtained by telephone prior to initiating contact with the sample member. As a last resort, cases were also assigned to field staff for an in-person interview.

As shown in table 34, the majority of those who responded (74 percent) did so during the in-school Survey or Makeup Day. Approximately 20 percent participated as a result of the telephone interview follow-up. Just over 5 percent were interviewed by a field interviewer and less than one half of one percent completed a mail questionnaire.

Table 34. Overall yield, by method of data collection (unweighted percentages): 2004

Method	Number of responses	Percent of total response
Total responses	14,989	100.00
In school	11,125	74.21
Mail	43	0.29
Telephone	3,024	20.17
Field	797	5.33

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), “First Follow-up, 2004.”

4.2.3 Data Collection for Transcripts and Course Offerings

Transcripts were collected from sample members at the end of 2004 and early in 2005, a minimum of 6 months after most students had graduated from high school. Transcripts were collected from the students’ base-year school. However, if it was learned during the first follow-up data collection that the sample member had transferred, transcripts were collected from two schools: the base-year school and the last known school of attendance. For students who were added to the study during the spring term of their senior year (known as “freshened” students), transcripts were only collected from their senior-year school. Transcripts were collected for regular graduates, as well as dropouts, students still in high school, early graduates, and students who were homeschooled after their sophomore year.

Transcripts were collected for all sample members who participated in at least one of the first two student interviews: the base year or the first follow-up. These sample members include base-year respondents who were first follow-up nonrespondents and base-year nonrespondents who were first follow-up respondents. Thus, sample members who were dropouts, freshened

sample members, transfer students, homeschooled students, and early graduates were included if they were respondents in either of the first two student interviews. Transcripts were also requested for students who could not participate in either of the interviews because of a severe physical disability, a mental disability, or a language barrier. A total of approximately 1,500 base-year and transfer schools responded positively to the transcript request by providing transcript data for ELS:2002 sample members. Ninety-one percent (weighted) of the ELS:2002 student sample have transcript information (about 14,900 out of 16,400).

Records were necessarily incomplete for sample members who had dropped out of school, had fallen behind their cohort's modal progression sequence, or were enrolled in a special education program requiring or allowing more than 12 years of schooling. Eighty-six percent of transcript respondents have 4 complete years of high school transcript information.

4.2.3.1 Transcript Data Collection Materials

The development of data collection materials and procedures was informed by the NELS:88 high school transcript study, the National Assessment of Educational Progress high school transcript study, and the field test for ELS:2002 transcript data collection. Data collection materials were mailed to schools beginning in December 2004. The materials were sent to the ELS:2002 school coordinator at all schools that participated in ELS:2002. If the school was new to the study (e.g., a school attended by a sample member who transferred out of his or her base-year school), the materials were sent to the principal. The materials guided school personnel in the preparation of transcripts and related documents. Each school was asked to provide basic enrollment, testing, and coursetaking information for each student, as well as information about the school's grading and graduation policies and requirements. The information requested included the following:

- Student-level information, including
  - type of diploma awarded (e.g., standard, honors, or General Educational Development certification);
  - date diploma awarded;
  - date student left school;
  - reason student left school (e.g., graduated or transferred);
  - cumulative GPA; and
  - test scores for the Preliminary Scholastic Aptitude Test, Scholastic Assessment Test, ACT,[51] and Advanced Placement tests.
- Coursetaking histories for grades 9 through 12,[52] including
  - course title and number;
  - year, grade level, and term course taken;

[51] Formerly called the American College Testing Assessment.

[52] Schools were also encouraged to provide information about coursetaking immediately prior to 9th grade, especially algebra or geometry courses. These courses appear on the course-level file but are not included in any of the composite measures on the student-level file.

  - number of credits earned; and
  - grade assigned.
- School-level information, including
  - grade scale;
  - course grade weighting system used, if any;
  - availability of student-level information;
  - GPA formula;
  - Carnegie unit conversion information;
  - term system;
  - course catalogs (if not collected previously);
  - types of diplomas granted; and
  - credits required for different types of diplomas.

The data collection materials requested from school personnel also included the following: cover letter, instructions for preparing transcripts, student transcript checklist, transcript cover sheet, disclosure notices, value and uses of transcript research document, and signed consent forms (if the school required explicit consent).

The instructions for preparing student transcripts requested that photocopies or printouts of transcripts be prepared for the students listed on the Student Transcript Checklist. They also requested that the transcripts, when available, include coursetaking histories for 9th through 12th grades. In the rare instances in which 9th-grade records were unavailable, the preparer was asked to submit photocopies or printouts of transcripts for the 10th through 12th grades.

4.2.3.2 Transcript Data Collection Procedures

From December 2004 through June 2005, survey materials were sent to over 2,000 schools. This group included schools that participated either in the base-year or first follow-up survey and transfer schools that were first contacted regarding ELS:2002 during transcript data collection. Transcripts were not requested from 10 base-year schools because they had refused to participate in the first follow-up survey. Additionally, transcripts were not requested from one base-year school that had no eligible students. Schools were paid $5 for each transcript.

Transcripts were requested for over 16,000 sample members. Included were sample members who were ineligible to participate in the base year or first follow-up because of a physical disability, a mental disability, or a language barrier. Ninety-five schools required explicit consent from sample members or their parents/guardians before releasing transcript information. Of the sample members who attended these schools, about a quarter provided signed release forms.

Two weeks after the survey materials were sent to the school, a follow-up postcard was sent as a reminder to complete the data collection forms and to send the requested materials to RTI. If after an additional week RTI had not received the materials from the school, assigned institutional contactors (ICs) began telephone prompting to request that the materials be sent as

soon as possible. Nonresponding schools contacted during the telephone prompting frequently requested remailing of the data collection materials. During telephone contacts, the ICs also identified any additional requirements the school had for releasing transcripts.

Telephone follow-up with schools continued through June 2005. Additional measures were implemented to ensure an adequate response rate. In June 2005, data collection materials were sent to schools that had not yet provided all of the requested transcripts. In addition, in-person visits to nonresponding schools were conducted during April through June 2005 to collect the requested materials or to assist the school transcript preparer in assembling the information. For efficiency, the schools were selected for in-person visits by their proximity to other schools. In-person visits were made only to schools that had not sent transcript materials for any requested sample members.

Collection of transcripts for dropouts and alternative completers was impeded in some cases by the frequency with which sample members transferred schools or dropped in and out of school, and inaccurate school records. Dropouts occasionally were enrolled in a school for too brief a period to accumulate a coursetaking record. Consequently, there is often little or no record of their destination school. However, the strategy of beginning by collecting transcripts from the school of origin (base-year school) maximizes the number of 2004 dropouts for whom there are at least complete 9th and 10th grade (fall 2000 to spring 2002) records.

4.2.3.3 High School Transcripts of Transfer Students

In addition to collecting data from base-year schools, transcript data were collected from the transfer schools of students who left their base-year high school. Transfer students were identified at several points in the ELS:2002 data collection process. These time periods included enrollment status update contacts in spring 2003, fall 2003, and spring 2004, as well as the first follow-up data collection in spring 2004.

4.2.3.4 Obtaining Permission for Collecting High School Transcripts

Because the Department of Education, under the Family Educational Rights and Privacy Act, has the right to obtain transcripts without prior consent for evaluation purposes, and because RTI informed parents, students, and school personnel of the transcript data collection as part of base-year and first follow-up data collection/consent activities, the first approach to collecting transcripts was a direct mail request to each school. When RTI contacted schools to prompt for return of the transcripts and answer any questions, it was also noted whether the schools had additional consent requirements before they would release student transcripts. Approximately 100 schools requested explicit consent. For sample members who attended these schools, RTI sent a letter and form to the students and their parent/guardian informing them that a signed consent form was required in order for the school to release the transcript to RTI. The consent letters explained that a parent's signature was required if the sample member was under age 18 and a sample member's signature was required for students 18 years of age or older.

After explicit consent forms granting permission to release the transcript were received, a second set of data collection materials was sent to each school requiring consent, including a list of students for whom signed consent was received and photocopies of the signed consent forms.

4.2.3.5 Course Catalog Data Collection

RTI began collecting course catalogs in the 2001–02 school year as part of base-year data collection activities and continued in the 2003–04 school year as part of first follow-up data collection activities. Course catalogs were requested for four school years covering 2000–04 from base-year schools and for two school years, 2002–03 and 2003–04, from transfer schools. In the first follow-up, course catalogs were requested from both base-year and transfer schools, for use in coding transcripts. However, only information for base-year schools appears on the course offering file. During the transcript request activities, schools were prompted for catalogs that had not yet been collected. If a school did not have a conventional catalog, then a course list, master teaching schedule, or any other form of information from which course offerings could be extracted was accepted. The course offerings response rate for base-year schools (the basis for the course offerings file) was 88 percent.

4.2.3.6 Definition of a Transcript Respondent

A sample member was considered a respondent in the ELS:2002 transcript data file if the following criteria were met:

- The sample member had at least one transcript sent from one school.
- The sample member had at least one complete course record for at least one grade (9th, 10th, 11th, or 12th).

4.3 Second Follow-up Data Collection Methods and Results

This section describes the data collection procedures and presents the data collection results for the ELS:2002 second follow-up survey. Section 4.3.1 details the data collection activities and procedures followed, including sample maintenance, tracing, respondent incentives, survey modes, and refusal conversion. Section 4.3.2 presents and discusses data collection outcomes and data quality indicators, including overall response rates, response rates across modes, response rates among key subgroups, interviewing effort, interview timing data, and quality assessment monitoring results.

4.3.1 Data Collection Activities and Procedures

This section describes the data collection activities and procedures followed in conducting the ELS:2002 second follow-up in 2006. The section is divided into five primary topics: pre-data collection tracing and sample maintenance activities, use of incentives, overview of data collection modes, intensive tracing efforts, and refusal conversion procedures. Procedures and activities under these five topics are described in this section, while the pertinent results and outcomes are presented in section 4.3.2. Maintaining data security is a requirement that pervades all tasks, including, of course, data collection. It is discussed in conjunction with the related topic of confidentiality protections associated with treatment of the analytic data in chapter 6, section 6.6.

Data collection for the 2006 round was significantly redesigned to include survey modes and procedures that were completely independent of the in-school orientation of the first follow-up survey. Almost all of the young adults in the 2006 sample transitioned from high school to postsecondary education, the workforce, or the military between the first and second follow-up data collection periods. The 2006 data collection procedures focused on two critical elements for

reaching and enlisting these sample members: tracing and sample maintenance activities and multiple modes of data collection (web self-administration, in-person or telephone computer-assisted interviewer administration). The tracing and sample maintenance activities involved a variety of techniques to maintain current contact information for sample members and continue their identification with ELS:2002 for the second follow-up round. Offering multiple modes of data collection maximized the opportunity for sample members to participate in the ELS:2002 second follow-up. Combined, these two important features of the 2006 round of data collection plan were designed to include a very high percentage of this mobile population in the survey. In addition to drawing on experiences from conducting the second follow-up field test data collection, the second follow-up procedures drew on the experiences of other education surveys with similar populations, including Baccalaureate and Beyond 2003 (B&B:03), the National Postsecondary Student Aid Study of 2004 (NPSAS:04), and the 2004/2006 Beginning Postsecondary Students Longitudinal Study (BPS:04/06).

In addition, key characteristics of second follow-up sample members were considered in developing plans to achieve a high response rate. While the overwhelming majority of F2 sample members participated in both the base-year (BY) and F1 waves, 14 percent participated in only one of the previous two rounds. The data collection procedures anticipated that those who did not participate in 2004 would likely be especially difficult to include 2 years later. Furthermore, the sample included high school dropouts, who proved more difficult to locate and include in the 2004 data collection. As a result, second follow-up data collection procedures included features to help maximize participation among these special subpopulations in the sample.

4.3.1.1 Pre-Data Collection Tracing and Sample Maintenance Activities

An important aspect of the ELS:2002 second follow-up (2006) data collection was that high schools were no longer involved in providing assistance with locating sample members. High schools had played a central role in facilitating contacts and interviews with participants in both the BY and F1 rounds. In addition, as the young adults in the sample transitioned from high school to different educational and work pursuits they moved away from their previous homes. For this reason, a more extensive set of tracing and sampling maintenance techniques was warranted for the second follow-up, including the following:

- batch tracing services for updated address information and telephone numbers;
- updated locating information obtained from student federal financial aid applications (FAFSA);
- direct contact with sample members and their parents via mail, telephone, or Internet;
- intensive tracing efforts by centralized tracing specialists;
- intensive tracing efforts by field locating specialists in local areas; and
- tracing students through postsecondary schools applied to or attended, as specified in the 2004 interview.

Another important tool in this process was to continually build on the positive relations ELS:2002 staff have established and maintained with sample members. All contacts with sample members were designed to maintain cordial and respectful relations with the young adults in the sample and their parents. This section describes the pre-data collection tracing and sample

maintenance activities implemented prior to the start of 2006 data collection. Section 4.3.1.4 describes intensive tracing activities conducted during the second follow-up data collection period. For reference, table 35 outlines the complete schedule of all tracing and sample maintenance activities for the 2006 sample.

Table 35. Tracing and sample maintenance activities for the ELS:2002 sample: 2004–08

Date	Activity
December 2004	NCOA and Telematch batch tracing of sample members and parents
October 2005	NCOA, Telematch, and CPS batch tracing of sample members and NCOA and Telematch batch tracing of parents
November 2005	Direct mailing to sample members/parents with toll-free line, e-mail, and website request to update contact information
January 2006	Pre-data collection intensive tracing of sample members without valid current contact information
February–September 2006	Intensive centralized tracing and field locating during F2 data collection
April 2007	NCOA, Telematch, and CPS batch tracing of sample members and NCOA and Telematch batch tracing of parents
April 2007	Direct mailing to sample members/parents with toll-free line, e-mail, and website response to update contact information
April 2008	NCOA, Telematch, and CPS batch tracing of sample members and NCOA and Telematch batch tracing of parents
April 2008	Direct mailing to sample members/parents with toll-free line, e-mail, and website response to update contact information

NOTE: NCOA = National Change of Address. CPS = Central Processing System, Free Application for Federal Student Aid (FAFSA), U.S. Department of Education federal educational loan application database.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Batch tracing activities. Pre-data collection tracing and sample maintenance efforts for the 2006 sample began in December 2004 with the updating of sample members' contact information through batch tracing services. (To minimize costs, the full-scale sample was combined with the field test sample in these activities.) Batch tracing represents a cost-effective method of updating addresses and telephone numbers for the young adults in the sample and their parents. Two batch tracing services, National Change of Address (NCOA) and Telematch, were used to update the address and telephone information sample members provided in previous rounds. NCOA is a database consisting of change of address information submitted to the U.S. Postal Service. Matching ELS:2002 sample members' addresses against the NCOA database was useful for providing address updates, especially for those sample members who had recently moved. The Telematch service involves a database of over 130 million residential listings, including over 3 million unlisted numbers that have recently been assigned. Telematch was used to confirm or update the telephone number for each sample member matched to their most current known address. These two services are most effective when used in this sequence, because the updated addresses from NCOA can be matched to sample members' updated telephone numbers, when applicable. A total of about 16,400[53] second follow-up cases were submitted to these batch tracing services in December 2004. All information returned from these batch tracing services was then updated in the sample locator database.

[53] Readers are reminded that second follow-up sample sizes for subgroups are approximate. There is no public-use data file for the second follow-up. Exact sample sizes from restricted-use data files cannot be published unless the data are perturbed in some way. The perturbation approach taken here was to round the exact sample sizes of cells to tens (if less than four digits) or hundreds (if four or five digits). This convention has been followed even for methodological tables containing information excluded from the ECB.

The next set of tracing and sample maintenance activities for 2006 sample members began in October 2005. At this point, three batch tracing services were used to update sample member contact information—NCOA, Telematch, and the U.S. Department of Education's Central Processing System (CPS) for federal financial aid applications. The CPS search was added to confirm or update contact information for those sample members who had submitted the FAFSA for one or more postsecondary institutions. The CPS database was another useful source of information for locating a significant number of sample members, because 74 percent of 2006 sample members had attended postsecondary schooling and an estimated 70 percent of these attendees completed a FAFSA.

A total of approximately 16,200 second follow-up sample members had sufficient information to send to NCOA and Telematch. Table 36, table 37, and table 38 provide the results for NCOA, Telematch, and CPS batch tracing, respectively. First, the NCOA database provided new address information for 2,300 sample members (14 percent). Approximately 210 cases (1 percent) were identified as no longer valid, but no new information was available. The most common result of this tracing service was reflected in the 13,800 cases for which no match was obtained in the NCOA database. This result could indicate that either these cases had current address information that had not recently changed or that they simply were not included in the NCOA database. Additionally, about 70 second follow-up sample members did not have sufficient current address information to be included in the NCOA batch tracing. Another 50 sample members had a finalized status such as final refusal or ineligible.

Once new information from the results of the NCOA batch tracing were updated in the locator database, about 16,300 cases were submitted to the Telematch batch service. Among these sample members, about 9,000 (56 percent) did not have a matching record in the Telematch database. Of the remaining 46 percent of cases, the majority (42 percent) resulted in confirmation of the telephone number on record. The other 500 cases (3 percent) did produce a match to a new telephone number.

The final step in this sequence of batch tracing was to match the contact information for sample members against the U.S. Department of Education's CPS FAFSA application database. Because this process included both address and telephone information, the results were somewhat more complicated than the NCOA and Telematch services. For about 3,700 sample members (23 percent), the existing contact information was confirmed in the CPS database. New information was obtained for 1,300 sample members (8 percent), which were nearly evenly divided among new address information only, new telephone number only, and both new address and telephone, as indicated in table 38. A total of about 7,200 cases (44 percent) sent for CPS matching produced no match in the database. The remaining approximately 4,100 cases did not have valid Social Security numbers and therefore could not be matched against the CPS database.

In anticipation of potentially contacting sample members' parents as part of 2006 data collection, address and telephone information was also sent to NCOA and Telematch for one "primary" parent of 16,000 sample members. The "primary" parent was selected by prioritizing all parents identified in prior rounds of ELS:2002 by their relation to the sample member, starting with mother. The remaining 320 sample members did not have sufficient information for one or more parents to be included in this batch tracing activity. Once again, the updated contact information obtained through these batch tracing activities was then used to update the second follow-up sample locator database in preparation for future contacts.

Table 36. NCOA batch tracing results for second follow-up sample members: 2006

Tracing outcome	Number of cases	Percent of cases
Total cases	16,400	100.0
Bad address—new information obtained	2,300	13.9
Bad address—no new information obtained	210	1.3
No match found in database	13,800	84.4
Insufficient address information to be included in batch	70	0.4
Finalized status of refusal or ineligible	50	0.3

NOTE: Detail may not sum to totals because of rounding. Case numbers have been rounded to tens or hundreds. NCOA = National Change of Address.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Table 37. Telematch batch tracing results for second follow-up sample members: 2006

Tracing outcome	Number of cases	Percent of cases
Total cases	16,400	100.0
Obtained new telephone number	500	3.0
Confirmed telephone number	6,800	41.5
No match found in database	9,000	55.5
Finalized status of refusal or ineligible	50	0.3

NOTE: Detail may not sum to totals because of rounding. Case numbers have been rounded to tens or hundreds.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Table 38. CPS batch tracing results for second follow-up sample members: 2006

Tracing outcome	Number of cases	Percent of cases
Total cases	16,400	100.0
New address and new phone number obtained	550	3.4
New address obtained	400	2.5
New phone number obtained	350	2.2
Confirmed existing address and phone number	3,700	22.6
No match found in database	7,200	44.1
Insufficient address information to be included in batch	4,100	25.4
Finalized status of refusal or ineligible	50	0.3

NOTE: Detail may not sum to totals because of rounding. Case numbers have been rounded to tens or hundreds. CPS = Central Processing System, U.S. Department of Education, Federal Loan Application Database.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Sample maintenance contacts. In addition to maintaining current contact information for ELS:2002 second follow-up (2006) sample members and their parents, batch tracing efforts were also important preparation for direct contact with sample members and their parents. Periodic contacts with sample members and their parents maximized sample members' connection to ELS:2002 over the duration of the study. Direct mailings to sample members and their parents provided an opportunity to thank sample members for their continued participation in ELS:2002, inform them about the next steps in the research, and request that they review and update their contact information currently on record.

Between the F1 and F2 survey rounds, a direct mailing to sample members and their parents was sent in November 2005. A total of approximately 16,300 sample members and at

least one of their parents had sufficient address information for sending this packet. Of these, about 360 sample member packets (2 percent) and 770 parent packets (5 percent) were returned with no forwarding address information available. Another 670 sample member packets (4 percent) and 310 parent packets (2 percent) did return forwarding address information so that these packets could be remailed.

This mailing included a cover letter thanking sample members or parents for their continued participation, notifying them of the next round of data collection, and requesting that they update their contact information in anticipation of data collection. The packet also included a contact information form for sample members or parents to provide updated name, address, telephone, and e-mail information. Separate letters and forms were sent to sample members and parents to appropriately present and capture the contact information. The most current information for sample members and their parents was preprinted on the contact information forms, so that respondents could easily indicate any corrections to this information. The forms provided spaces for both permanent and current contact information, which was useful for sample members who were away at school or in the military, but still considered their parents' home address as their permanent address. Sample members and parents were provided with instructions on the form to either confirm or correct their contact information.

To respond to the sample maintenance mailing, sample members or their parents were able to update their contact information in one of four ways:

1. Completing and returning the hardcopy contact information form in the return envelope provided.
2. Updating the contact information via the ELS:2002 website, which provided a link ("Update your Contact Information") to online forms similar to the paper contact information forms.
3. E-mailing updated information to the dedicated ELS:2002 e-mail address.
4. Providing updated information by calling the dedicated ELS:2002 toll-free line.

This sample maintenance activity was useful for both maintaining contact with sample members and obtaining updated information from those who had recently relocated to attend college, take a job, or serve in the military. In addition, because letters were sent to parents as well, a number of parents either returned the contact information forms, logged on to the website, or called the toll-free number to update their child's information.

Like most sample maintenance activities of this kind, overall response to the November 2005 mailing to sample members was limited. Overall, about 1,620 sample members (10 percent) and about 1,850 parents (11 percent) returned the hardcopy contact information forms. The slightly greater response from parents is consistent with prior rounds of ELS:2002, where parents were consistently helpful sources of information about their children. Additionally, approximately 940 sample members or their parents updated their contact information using the study website, for an overall total of about 4,400 updates. Of the 940 updates entered via the website, about 160 were completed prior to the start of data collection on January 25, 2006, and 780 were completed once data collection began. Only a small number of those who responded to the sample maintenance mailing used e-mail or voicemail messages. The web option may have served to increase the overall response to the sample maintenance mailing by providing a mode with which sample members would be comfortable. This outcome cannot be directly assessed,

however, because use of the website may have simply substituted for returning the contact information forms among responding sample members.

Pre-data collection intensive tracing. The batch tracing activities and sample maintenance mailing conducted in November–December 2005 identified a set of 90 second follow-up sample members for whom no current contact information was available. These cases were sent to the centralized Tracing Operations Unit (TOPS), a part of RTI's Call Center Services, for intensive tracing. Among these cases, tracing specialists obtained new information for 40 cases (48 percent). This new information was then used to update the 2006 sample locator database for these cases. One additional case was identified as being out of scope for the 2006 data collection through tracing efforts. For the remaining 40 cases (51 percent), TOPS was unsuccessful in obtaining any updated contact information. These cases were among the first to undergo further intensive tracing efforts during the data collection period.

4.3.1.2 Use of Incentives

Incentive payments to respondents was one feature of the data collection plan for the ELS:2002 2006 study. Even following the best survey practices, collecting data from some respondent populations is difficult and expensive, making it almost impossible to obtain desired response rates without incentives. The results of the 2003 field test experiments and the success of the 2004 round of data collection provided evidence of the value of respondent incentives in achieving high response rates (see Ingels et al. 2005b, appendix J). A number of important factors were considered in developing and implementing the incentive plan:

- Almost all first follow-up sample members received an incentive, including both those who participated in school and those who participated outside of school. Paying incentives to almost all first follow-up participants raised the expectation among the sample cohort that they would receive payment again for participating in the 2006 round.
- Between the F1 and F2 surveys, the ELS:2002 sample cohort became further dispersed. In both the 2004 main study and second follow-up field test (2005) providing incentives was effective in making contact with sample members who were difficult to reach.
- Offering incentive payments can actually reduce data collection costs by limiting the effort required to pursue sample members who are difficult to locate or those who are initially unwilling to participate. Significant cost savings are gained by reduced computer-assisted telephone interview (CATI) and computer-assisted personal interview (CAPI) follow-up efforts, including repeated contacting attempts, refusal conversion calls, and field interviewer travel.
- Although cell sizes for important analytic subgroups were satisfactory after the success of the 2004 data collection, significant attrition among these subgroups was a threat to the analytic value of the second follow-up. The two most important subgroups that were offered higher incentives in the first and second follow-ups were high school dropouts and prior-wave nonrespondents. Paying differential incentives to both dropouts and first follow-up nonrespondents in 2006 was designed to ensure sufficient inclusion of these important subgroups.

The second follow-up incentive plan was designed to maximize respondent participation by meeting their expectations of compensation for their time and efforts, helping to locate widely dispersed sample members, and offering greater incentives to particular subgroups with limited representation in the sample. In addition, the incentive plan was generally similar to the 2004 plan and also incorporated elements of similar education studies, including NPSAS:04 and the BPS longitudinal study. In this way, the 2006 plan was as consistent as possible with both the prior round of ELS:2002 and other current education surveys of the young adult population.

The 2006 incentive plan was designed to address five key features of survey context:

1. First follow-up participation status—F1 respondent or F1 nonrespondent.
2. High school dropout status—identified in F1 as ever having dropped out or not.
3. Timing of participation—during the first 4 weeks of data collection or beyond this period.
4. Difficulty in contacting or enlisting cooperation with the sample member—meeting the criteria for difficult cases or not.
5. Partial prepayment of the incentive for sample members who had not participated after all other incentive conditions had been exhausted—completed prior to the final 8 weeks or beyond this period.

The first four of these five elements were approved by the U.S. Office of Management and Budget (OMB) and established prior to the start of the data collection period. The fifth element was implemented as a contingency during data collection based on discussions with and approval from OMB.

Because multiple criteria applied to many sample members, the incentive plan elements were combined to determine the appropriate payment level at each point of the study. In order to ensure that survey notification materials and interviewer statements matched respondents' expectations on how much they would be paid at each point in the data collection period, consistency was maintained across all points of contact with respondents regarding the amount of their incentive payments. This consistency was achieved initially and maintained throughout the study by using the same predetermined variables—dropout status, F1 participation status, difficult case status, and current date—in all study materials and computer programs to indicate the appropriate incentive amount. Materials included mailed letters and instructions and e-mail messages. Computer programs included web/CATI/CAPI scripts and instruments as well as the sample database. The same procedures followed in the 2006 round to ensure consistency had been used effectively in the 2004 data collection.

Table 39 summarizes the specific elements of the 2006 incentive plan. The regular or "base" incentive amount for all ELS:2002 sample members who had never been identified as dropouts and had participated in the F1 data collection was $20. For those sample members who participated in the base-year study but did not participate in 2004, the regular incentive was higher at $40. Likewise, those who had ever been identified as dropouts through the 2004 round were offered $40 as a base incentive.

Table 39. Second follow-up full scale respondent incentive plan: 2006

Respondent type	Regular incentive	Early completion	Difficult case	Final difficult ($10 prepaid)
F1 nonrespondent	$40	$50	$50	$60
Ever dropout	40	50	50	60
F1 respondent, nondropout	20	30	30	40

NOTE: F1 = First follow-up.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

In order to encourage sample members to participate early in the data collection period, either through web self-administration or by calling the toll-free number to complete a CATI interview, those who completed the survey (by either mode) prior to the start of outbound CATI calling were paid an additional $10 on top of the regular incentive. The early incentive period ran from the opening day of data collection on January 25, 2006, through February 19, 2006, when outbound calling began. This element was designed to offer the most responsive sample members a bonus for participating prior to when more intensive data collection procedures were implemented.

A further addition to the incentive payment plan was to offer an additional $10 over the regular amount to those sample members who proved extremely difficult to contact or enlist in the study during the course of the 2006 data collection period. This increase was implemented independently of each sample member's high school completion status or F1 participation status. The criteria for the "difficult" status increase included the following:

- more than 20 calls were made to contact the sample member without completing an interview;
- sample member refused to participate during an initial contact;
- others refused multiple times on behalf of the sample member;
- sample member could not be located through any of the telephone numbers previously provided, so the case was submitted for intensive tracing;
- case was sent to a field interviewer for tracing; or
- sample member had still not completed the interview as of June 15, 2006.

Once a case met one (or more) of these criteria, all computer programs and databases were updated with the higher incentive amount.

The preceding elements of the respondent incentive plan were all implemented at the beginning of the 2006 data collection period. On July 6, 2006, one final revision to the incentive plan was implemented for the final 8 weeks (or about 2 months) of data collection. All sample members who had not yet completed the survey were sent an express mail package with an additional $10 check as a prepayment of the full incentive amount. The remainder of the incentive was payable upon completion of the survey. If mailed packages did not reach the intended sample members and at least one alternative address was available in the sample members' records, data collection staff remailed the $10 prepaid check to these sample members. The purpose of the prepaid incentive was to assure remaining sample members that NCES and RTI were serious about obtaining their participation in the survey and compensating them for completing the survey. A total of 3,200 packages with the prepaid incentive check were mailed.

Another 10 sample members who had not yet completed the F2 interview did not have a current, valid address to be included in this mailing.

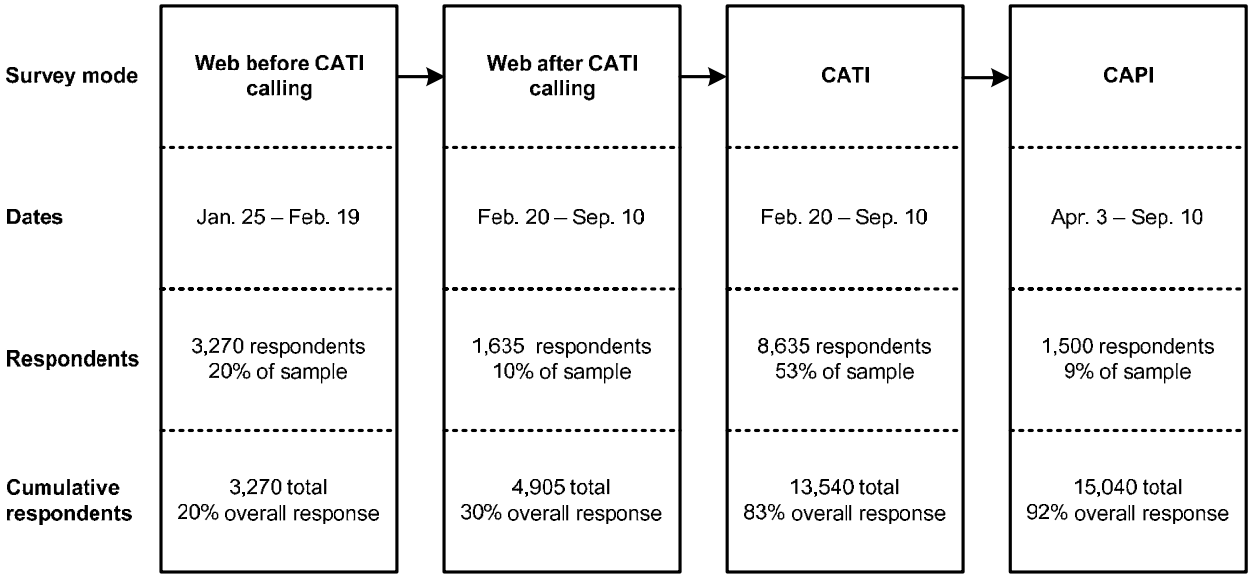
Throughout the 2006 data collection period, all incentive payments were provided in the form of checks. The data file for incentive payments was created at the beginning of each week and the incentive checks and thank you letters were mailed to participants at the address indicated during the last section of the interview. Because address information was occasionally incomplete or inaccurate, data collection staff investigated returned incentive checks to find an accurate mailing address so that these checks could be remailed. Subsequent sections present the counts and percentages for each type of incentive payment paid over the course of the 2006 data collection.

4.3.1.3 Overview of Data Collection Modes—Web, CATI, and CAPI

Multiple modes of data collection was a second important feature of the 2006 data collection. Figure 3 outlines the data collection schedule and targets for each survey mode—web computerized self-administered questionnaire (web CSAQ), CATI, and CAPI. Offering 2006 sample members web, CATI, and CAPI modes provided three viable, complementary modes for gaining cooperation. Providing these multiple modes also eliminated having to devise and administer a hardcopy version of the survey instrument. Because it would have been nearly impossible to anticipate the appropriate set of questions for each individual sample member, hardcopy questionnaires would have likely placed an undue burden on respondents to correctly navigate through the instrument. Furthermore, hardcopy questionnaires would not be directly comparable to the computer-assisted versions, as a number of 2006 survey items relied on computer logic, preloaded data, and help features that would not have been available in hardcopy format.

Offering sample members the self-administered web-based survey option in addition to CATI and CAPI survey modes was a major advance in the 2006 data collection. Web self-administration was viewed as a viable data collection mode based on data that the 19- and 20-year-olds who comprised the sample would have relatively high rates of Internet access and usage. In the second follow-up field test, nearly 28 percent of sample members participated via the Web. For 2006 data collection, the expectation was that appropriate procedures to facilitate and encourage participation via the ELS:2002 website would result in about 30 percent of sample members completing via web self-administration mode. The web-based mode was expected to make a substantial contribution to overall response to the 2006 data collection and, at the same time, conserve survey resources for CATI and CAPI follow-up with remaining sample members.

Figure 3. Data collection flow across survey mode for the second follow-up full scale study: 2006



NOTE: CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Data collection in 2006 began as scheduled on January 25, 2006. For the first 4 weeks, only web and call-in data collection was made available to sample members. To notify sample members about the start of data collection, all sample members and parent(s) were sent a packet that included the following:

- a cover letter explaining the goals of the ELS:2002 study;
- directions for accessing the ELS:2002 website;
- instructions for completing the interview on the Web or calling in to complete by phone;
- a unique user identification and password for each sample member to access the web interview (this was only included in the sample member letter);
- a toll-free help desk phone number to call for assistance with web self-administration or completing the interview by phone;
- a toll-free number and e-mail address for any general questions about the ELS:2002 second follow-up; and
- an informational brochure describing the ELS:2002 study.

To communicate the importance of each sample member's participation, the lead letter was sent on Department of Education letterhead and signed by the NCES project officer for ELS:2002.

At the start of data collection (January 2006) a total of 16,100 lead letter packets were sent to sample members and 15,800 to parents of sample members. Packets were not mailed to cases if no usable address existed or if a finalized status had been determined as a result of pre-data collection contacts with sample members or parents. Among those mailed, 350 sample

member packets and 130 parent packets were remailed with updated contact information returned from the original mailing. Another 270 sample members' packets and 490 parent packets were returned without forwarding address information. In addition to informing most sample members and parents about the start of data collection, the lead letter mailing identified a small set of cases for whom some or all existing contact information was no longer valid. Most of these cases had limited contact information in the database and the address used for the lead letter mailing proved to be no longer valid.

Web self-administration. Offering sample members the option of completing a self-administered interview via the ELS:2002 website was a major enhancement to second follow-up data collection. The web-based survey mode provided several important advantages in collecting data from the population of young adults who were included in the 2006 sample:

- A high percentage of sample members would be familiar with using the Web, and many were likely to have already completed other web surveys.
- Sample members could complete the web interview at any location where they had Internet access, which was convenient for young adults who are generally active and mobile.
- Sample members could complete the web interview at any time that was convenient for them, which was particularly advantageous for those who were busy with work or school in the afternoon and evening hours.
- Web self-administration allowed respondents to complete the interview at their own pace, which is attractive both to those who can move quickly through the instrument and those who need to take more time.
- Sample members who preferred not to discuss certain aspects of their high school experiences since the last interview could achieve greater privacy through using web self-administration.
- Because a web-enabled survey system was used to program the 2006 instrument, the web self-administered interview could be presented in a way that was virtually indistinguishable from CATI and CAPI modes.
- The web data collection mode was relatively cost-effective because it required a small support staff compared to the larger interviewing staffs required for CATI and CAPI data collection.
- Like CATI and CAPI interviews, web surveys provided faster access to data files than hardcopy questionnaires would have allowed, so this mode facilitated more timely review of survey data early during the fielding period.

A key assumption in implementing a web self-administration mode for the second follow-up was that a high percentage of sample members would have access to the Internet and be familiar with using web-based applications. Experience in the second follow-up field test indicated that the majority of ELS sample members had access to and familiarity with using the Internet. As a result, we expected a majority of second follow-up sample members to be comfortable using the Internet. Based on assumptions about sample members' access to and use of the Web, a total of 30 percent of sample members, or 4,900, were projected to complete the survey via web self-administration (see figure 3). It was expected that two-thirds of web

responders, or 20 percent of the sample, would respond during the early completion period when the early completion incentive was in effect. Another 10 percent of the sample was expected to complete web self-administration during the remainder of the data collection period.

Web data collection procedures. To facilitate web self-administration, a secure, dedicated ELS:2002 website was hosted on the NCES server. The ELS:2002 website could be used by sample members both to complete the survey and also to gain more information about the study. Respondents simply logged onto the website, clicked on a link labeled "Login to the Interview," and then entered the study identification number and assigned password to begin the interview. A "Frequently Asked Questions" ("FAQs") link also provided information about the study procedures and instructions for completing the web interview, so that respondents could obtain immediate help with any survey completion issues. Additional background information was also provided via the link "About ELS:2002." Through the course of the study, information on the website was added or revised to communicate any updates on data collection procedures and study timeline to sample members.

The primary strategy for conducting successful web-based data collection was to make self-administration as easy as possible for second follow-up sample members when they went to the website. To avoid technical problems, the web-enabled survey system was designed to function appropriately in a wide range of computing environments, including different web browsers, different Internet connections, and different computer settings. The login procedures were fairly simple and clearly explained in the lead letter mailing to sample members. Each screen of the instrument was designed so that the response task was clear. Special instructions were available at the click of a button to guide respondents through potentially problematic screens or to provide definitions of technical terms used in items. Although web help desk staff were available to assist respondents who had difficulties starting or completing the interview, development and testing of the web interview were designed to minimize these situations. When needed, web help desk staff were available through the toll-free ELS:2002 telephone line to provide technical assistance to respondents with computer, Internet, or survey issues.

All second follow-up sample members were initially treated as potential web respondents in the 2006 case management system (CMS). When sample members completed the web survey, this information was transmitted to the CMS. Once this information was captured in the CMS, further data collection contacts to these sample members was discontinued and the address information provided by participants was used to mail incentive checks. As detailed in section 4.3.1.2, the most responsive sample members who completed the web interview (or called in to complete a CATI interview) during the first 4 weeks were offered an early completion bonus. The expectation was that web response would be quite high during the first 4 weeks of data collection as a result of the higher incentive, and then taper off significantly in the ensuing weeks. Web self-administration was available to sample members throughout the entire 2006 data collection period. The web mode was therefore supported and encouraged over the course of data collection by CATI and CAPI interviewing staff and direct reminder contacts. Section 4.3.2.1 presents the results for web self-administration.

Web help desk staffing, training, and procedures. Even though the web-based survey protocol and instrument were designed to be easily completed, web help desk staff were hired and trained so that they would be available to assist sample members in completing the web interview. Initial training for help desk staff involved 20 hours total from January 20 through January 22, 2006. During or following training, all web help desk staff were required to

successfully complete certification requirements for both help-desk procedures and CATI interviewing. Help desk staff were trained to support web self-administration, call selected sample members to encourage early participation, and contact and interview sample members once outbound CATI interviewing began. Newly hired staff also had to complete basic RTI interviewer training prior to ELS:2002 web help desk training.

The goal of the web help desk training program was to provide staff with the opportunity to familiarize themselves with the study goals, the specific procedures, and the survey instrument, as well as the technical requirements and procedures for web self-administration. Key information on the purpose and goals of ELS:2002 and specific help desk and interviewing procedures were compiled in a manual for help desk staff to reference. All help desk staff were trained on how to address common issues or concerns of web respondents, both procedural and technical. The training provided help desk staff with technical information about web-based data collection so that they would be able to address respondents' technical questions. Technical information was summarized in an appendix to the web help desk/CATI interviewing manual provided to all help desk staff. This appendix could then be used by help desk staff to diagnose and resolve technical problems. All training topics were reinforced by group discussion and interaction, trainer demonstrations, and class practice sessions. Role-playing and individual practice were also important elements of the training. At the end of training, web help desk staff were certified for 2006 data collection by completing tests, exercises, and practice, including the following:

- homework exercise on knowledge of the ELS:2002 study;
- verbal test on responding to frequently asked questions from sample members;
- verbal test on pronouncing key terms featured in the interview;
- two complete practice interviews;
- exercise on handling different help desk scenarios;
- coding exercise for postsecondary fields of study, postsecondary institutions, and occupations; and
- coding exercise on case outcomes from inbound and outbound calls.

Completing these activities ensured that web help desk staff were well prepared to assist sample members to complete the web self-interview, administer CATI interviews using best practices, and persuade sample members to finish either the web or CATI interview.

Once data collection began on January 25, 2006, help desk staff were available to take incoming calls and respond to e-mails from sample members 7 days a week. Help desk hours were Monday through Thursday from 9 a.m. to 11 p.m. Eastern time, Friday from 9 a.m. to 9 p.m. Eastern time, Saturday from 10 a.m. to 6 p.m. Eastern time, and Sunday from 1:30 p.m. to 10:30 p.m. Eastern time. Help desk staff monitored the toll-free call-in line for calls and voice mail messages from sample members. Each morning, help desk staff checked the ELS:2002 voice mailbox for messages left by sample members after operating hours. E-mail messages were handled similarly. Help desk staff regularly checked the inbox for e-mail messages from sample members during operating hours and checked each morning for e-mails sent after operating hours. Help desk staff primarily communicated with sample members by telephone, but also responded to e-mails as appropriate. Communication with sample members was predominantly

in English, but the help desk staff included a sufficient number of bilingual interviewers to communicate with sample members (or parents) in Spanish, as necessary.

Additional web data collection activities. Beyond offering an early completion incentive to second follow-up sample members who completed the interview during the first 4 weeks of data collection, a number of additional follow-up activities were undertaken to encourage web participation in both the early completion period and beyond. During the early completion period, help desk staff conducted prompting calls to F1 nonrespondents in the sample to encourage them to participate. Because prior-round nonrespondents were likely to be difficult to contact, the prompting calls were intended to inform these sample members about the start of data collection and motivate them to participate early, either by web self-administration or CATI interviews. A secondary purpose of the prompting calls was to identify sample members who are unable or unwilling to complete the survey by web, so that help desk staff could offer to complete a CATI. Prompting calls for this subsample began on February 7, 2006, or just under 2 weeks after the start of data collection. The prompting protocol included up to three calls to F1 nonrespondents in the sample or until the interview was completed. Help desk staff communicated a brief statement on the purpose of the call to live respondents or on answering machine messages.

A second set of prompting activities begun during the early completion period was periodic mail and e-mail reminders to sample members who had not yet participated in the 2006 round. The first e-mail reminder was sent on February 3, 2006, about 10 days after the start of data collection. A first reminder letter was sent to sample members only several days later on February 8. Two additional e-mail reminders sent during the early completion period were an e-mail to parents on February 10 and an e-mail to sample members on February 15. The February 15 e-mail included a reminder to sample members that only a few days remained in the early bonus period. Additional e-mail and mail reminders were sent to remaining sample members throughout data collection. While these reminders were intended to encourage sample members to participate by any mode available at that time, reminder e-mails generally produced a spike in web self-administration over the next few days following each reminder. Table 40 lists all of the reminder e-mail messages and mailed letters sent to sample members throughout the course of data collection. These reminders were designed to promote sample member interest by varying the approach and focus of each message. For example, the July 6 reminder letter and July 7 e-mail message to sample members and the July 7 e-mail message to parents were used to inform sample members and their parents of the $10 prepaid incentive sent to them. Later e-mail messages in August and September emphasized the limited time remaining to participate. In addition, e-mails reminders were only sent to sample members who had not previously refused to participate. The e-mails and letters were also spread out sufficiently over the course of the data collection period so that sample members and parents had sufficient time to respond and did not feel inundated with contact attempts.

Table 40. E-mail reminder messages and mailed letter reminders to sample members and parents: 2006

Sample members	Parents of sample members
February 3—First reminder e-mail	February 10—First reminder e-mail
February 8—First reminder letter	March 23—Second reminder e-mail
February 15—Second reminder e-mail	April 14—Third reminder e-mail
March 10—Third reminder e-mail	May 18—Fourth reminder e-mail
March 31—Fourth reminder e-mail	June 22—Fifth reminder e-mail
April 28—Fifth reminder e-mail	July 7—Sixth reminder e-mail
June 9—Sixth reminder e-mail	July 20—Seventh reminder e-mail
June 30—Seventh reminder e-mail	August 10—Eighth reminder e-mail
June 30—First refusal reminder e-mail to sample	August 31—Ninth and final reminder e-mail
July 6—Second reminder letter	
July 7—Eighth reminder e-mail	
July 9—Second refusal reminder e-mail	
July 20—Ninth reminder e-mail	
July 27—Third reminder letter	
August 4—Tenth reminder e-mail	
August 14—Eleventh reminder e-mail	
August 23—Twelfth reminder e-mail	
August 31—Thirteenth reminder e-mail	
September 6—Fourteenth and final reminder e-mail	

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

CATI data collection. After the initial 4 weeks in which only web-based self-administration and call-in CATI interviewing were offered to sample members, outbound CATI data collection efforts were necessary to include more sample members in the 2006 round. The primary purpose of CATI data collection was to complete telephone interviews with sample members when contacted, or to set up an appointment to complete the interview. Outbound calls also served as reminders for some sample members who preferred to complete the survey via the Web. In this way, CATI interviewers often acted as proactive web help desk staff by offering assistance to sample members to complete the web survey. For example, CATI interviewers could provide sample members with their user IDs and passwords for logging in to the web survey, if they did not have their lead letter materials. For more technical problems, CATI interviewers who were not trained on help desk activities could quickly connect sample members to trained help desk staff for assistance. Overall, the expectation was that over half of all second follow-up sample members (53 percent) would complete the 2006 survey via CATI mode and another 10 percent would complete the survey via the Web after outbound CATI calling began.

The projections for CATI data collection anticipated completing interviews at a significantly higher rate in the first 3 months versus the last 3 months of the outbound CATI calling period. Still, the distribution was not expected to be as skewed as web data collection, because sample members were offered the incentive bonus to complete the interview prior to outbound CATI calls and CATI participation depended on contacting sample members by telephone at convenient times. Some sample members were easier to reach during the academic year of their postsecondary institutions, others were easier to reach in the summer months, and others were equally reachable (or unreachable) across the entire data collection period. In addition, many sample members were able to participate within the first few telephone contact attempts, while others required multiple callbacks to reach them at a good time to complete the

interview. The training schedule and staffing plan considered these factors to ensure that an appropriate level of resources were available for CATI interviewing over the course of the study. Section 4.3.2.1 presents the results for CATI interviewing.

CATI data collection procedures. All second follow-up sample members who had not yet completed web self-administration when outbound CATI data collection began were treated as part of the CATI sample in the CMS and managed via the CATI-CMS. The CATI-CMS facilitated sample management activities that included the following:

- providing sample members' and parents' contact information to interviewers;
- allowing scheduling of appointments to contact sample members at a specific time;
- automatically scheduling callback attempts for unsuccessful call outcomes such as ring-no answer, busy signal, and answering machine;
- appropriately coding sample members who were unable or unwilling to participate via CATI; and
- allowing for recording of relevant notes after each call attempt.

Cases that did not have any valid telephone numbers at the start of data collection were assigned to tracing in CATI-CMS, if they had not already been traced, or held for field tracing and data collection, if prior tracing had been unsuccessful. When sample members completed the web survey or otherwise reached a final status, this information was updated in CATI-CMS so that interviewers would discontinue calling these cases. Notes from each successive call attempt provided a "case history" that could be used by interviewers to make appropriate callback attempts and otherwise develop an effective strategy for reaching each sample member. Overall, CATI-CMS was the primary tool for facilitating efficient sample management for CATI data collection, tracing activities, and potential CAPI activities.

RTI's Call Center was open to receive and initiate calls Monday through Thursday from 9 a.m. to 11 p.m. Eastern time, Friday from 9 a.m. to 9 p.m. Eastern time, Saturday from 10 a.m. to 6 p.m. Eastern time, and Sunday from 1:30 p.m. to 10:30 p.m. Eastern time. Interviewers worked shifts to cover weekday, night, and weekend hours to ensure that both outbound and inbound calls from sample members were covered. Scheduling also ensured that outbound calls could be made at appropriate times across time zones and at times when sample members were likely to be reachable. Extended weekday hours until 12 midnight Monday through Thursday and 11 p.m. Sunday were also implemented during part of the data collection period to reach households in the Mountain and Pacific time zones.

When CATI interviewers contacted sample members, the CATI-CMS script prompted them to confirm the person's identity. The primary confirmation question (in addition to verifying the person's name) asked the sample member if he or she had attended the base year high school at any time in the past 5 years. If sample members did not verify attending the specified high school within the past 5 years, a follow-up question asked the person to provide his or her birthdate. If the birth date given matched, or very nearly matched, the birth date in the ELS:2002 database for the sample member, then the sample member's identity was confirmed. For a few cases, neither confirmation question was affirmed. These cases were identified as "problem" cases and reviewed by data collection staff to determine whether the person truly was not the sample member, or whether the high school and/or birth date in the database were

incorrect. CATI interviews were not completed until the interviewer confirmed the respondent's identity as the sample member. Once the person was confirmed, the interviewer read the informed consent script and then began the interview. All information collected in this part of the protocol was recorded by CATI-CMS for any further telephone contacts with the sample member.

The CATI instrument was virtually identical to the web self-interview. The only difference was that the CATI version provided an interviewer instruction on each screen to facilitate administration of each item. CATI interviewers adhered to standardized interviewing techniques and other best practices in administering the interview. Standardized interviewing procedures included reading questions and instructions exactly as scripted, probing insufficient answers in a nondirective way, recording respondents' answers exactly as given, and maintaining a neutral, nonjudgmental stance toward the substance of respondents' answers. This provided one potential advantage over web self-administration because CATI interviewers were able to provide appropriate guidance and probing whenever respondents were not completely clear on the response task for an item. CATI interviewers also functioned to reduce item nonresponse by ensuring that respondents attended to each question and probing respondents who were uncertain or hesitant about answering a question. All CATI interviewing procedures were conducted within the set of standardized interviewing techniques established prior to data collection and demonstrated in training. This not only ensured that interviewers were following appropriate interviewing procedures, but also minimized mode effects between self-administration and interviewer administration due to inappropriate guidance or feedback to respondents.

CATI training and staffing. A key to successful CATI data collection was hiring and training a quality, professional interviewing staff. The staffing plan was derived directly from the expectation for the interviewing hours needed each week to reach the goal for completed CATI interviews. Although all 2006 interviews were conducted in English, the CATI interviewing staff included a sufficient number of bilingual interviewers who could converse with parents or other relatives of sample members who communicate primarily in Spanish.

The CATI training sessions followed a similar agenda as the initial web help desk training in January. The key difference was that CATI trainings only included a basic overview of web help desk operations. This overview ensured that CATI interviewers were familiar with the kinds of technical assistance that the web help desk could provide and when to connect respondents to help desk staff. The reduced time spent on help desk issues in these CATI trainings was used to focus on effective strategies for enlisting cooperation among the young adults in the sample, as this was a central part of CATI interviewers' role in data collection. The enlisting cooperation sessions emphasized that the early stages of a call are most important for establishing trust and understanding among sample members and their parents. Interviewers were trained to effectively answer the common reasons for reluctance to participate. This training included developing a detailed knowledge of the purpose of ELS:2002 and why the participation of all sample members was vital to the success of the study. Small group discussions, refusal avoidance exercises, and other techniques were used to address these training issues. As with the help desk training, key information on the purpose and goals of ELS:2002 and specific interviewing procedures were compiled in a manual for CATI interviewers to reference.

At the completion of training, all telephone interviewers were required to be certified for data collection by successfully completing the tests, exercises, and practice, including the following:

- homework exercise on knowledge of the ELS:2002 study;
- verbal test on responding to frequently asked questions from sample members;
- verbal test on pronouncing key terms featured in the interview;
- two complete practice interviews;
- coding exercise for postsecondary fields of study, postsecondary institutions, and occupations; and
- coding exercise on case outcomes from inbound and outbound calls.

Completing these activities ensured that CATI interviewers were well prepared in both administering the interview using best practices and enlisting cooperation among sample members.

Quality control measures in CATI. In addition to training and certification procedures, a number of procedures were implemented to ensure and maintain data quality in CATI interviewing. Supervision and monitoring were maintained throughout data collection. Supervisors and monitors attended training alongside CATI staff so that they were familiar with the CATI interviewing procedures.

To directly assess the quality of telephone interviewing, RTI used two different monitoring protocols for CATI data collection. The first protocol, quality assessment (QA) monitoring, was designed specifically for U.S. Department of Education studies. This protocol focused directly on errors made by interviewers in administering individual survey items. While this monitoring protocol provided useful data on interviewer performance, its primary purpose was to track errors made for each survey question, to provide statistical evidence of data quality, and to ensure that the interviewing operations remained within acceptable statistical process control parameters. Each QA monitoring session involved a sample of up to 20 questions from an interview. In each session, monitors recorded the frequency of the following two types of errors:

1. Question administration errors that involve making changes in the question wording or response categories that significantly alters the intent of the question, or skipping the question inappropriately. Examples of question administration errors include the following:
   - changes to the reference period of the question;
   - changes to the "direction" of the question (e.g., from positive to negative wording);
   - changes to the frequency or duration of the question;
   - changes to a conditional term within the question; and
   - failure to read a question that should be asked.
2. Data entry errors such as keying an incorrect or inaccurate response to the question.

The data from QA sessions met multiple CATI data collection quality objectives, including identification of challenging questions, identification and reduction of any interviewer errors, and overall assessment of the quality of the data being collected. During the study,

ELS:2002 staff monitored daily results and charted the QA results on a weekly basis to detect any spikes in error rates and to make any necessary adjustments in data collection procedures. See section 4.3.2.4 for a presentation and discussion of the QA monitoring results for data collection.

A second CATI interviewer monitoring protocol focused more directly on interviewer behaviors related to performance. This standardized monitoring procedure was designed to record interviewer behaviors during the contacting and interviewing process that could then be used to provide interviewers with feedback to improve their performance. This monitoring served to reinforce appropriate behaviors in presenting the survey to respondents, enlisting respondent cooperation, delivering questions to respondents, probing inadequate responses, and maintaining professional and positive rapport with respondents. The interviewer-focused protocol for monitoring telephone interviewers focused on the following aspects of CATI interviewing:

- gaining respondent cooperation;
- delivering the introduction and informed consent;
- speech characteristics;
- reading skills;
- probing skills;
- professional skills in handling the interview situation;
- CATI navigation, coding, and data entry skills; and
- presentation skills.

Following individual monitoring sessions, CATI interviewers received timely feedback on any significant problems in enlisting cooperation or administering the interview. These monitoring session results were also compiled in a database to provide historical evidence of data collection quality for individual staff members, groups of staff members, and the entire project team.

Another important activity during CATI interviewing was holding regular quality circle (QC) meetings to ensure that procedures were being followed correctly and data quality was being maintained. The QC meetings provided a forum to focus on productivity goals, data quality, sample management, and related issues. The meetings also provided feedback to project managers on any issues that might require retraining of call center staff. After each meeting, managers summarized the issues discussed and provided resolutions of any problems to all call center staff in the form of QC memos. A total of 17 QC meetings were held during the course of CATI data collection, or about 1 every 2 weeks. To give all interviewers an opportunity to attend QC meetings regularly, separate meetings were held for day shift interviewers and night/ weekend interviewers.

A final important set of activities to ensure data quality in CATI operations was to follow statistical process quality control principles, including the following:

- measuring key indicators of data collection quality and quantity;
- reducing variation in the data collection process to maintain consistency;

- monitoring data collection indicators closely; and
- improving the process to maintain targets for both quality and quantity.

These principles were applied to three specific aspects of CATI operations: question administration and data entry, interviewer performance goals, and overall production goals. During data collection, these statistics were monitored daily and reported weekly so that CATI data collection could be regularly evaluated and any corrective steps taken quickly. For each indicator, we set acceptable ranges of variation against which results could be evaluated. The results were then presented in table or graphic form to facilitate quick identification of "out-of-control" conditions.

CAPI data collection. To reach sample members who had not yet participated by web or CATI modes, CAPI data collection commenced 8 weeks after the start of outbound CATI calling on April 17. Beginning field data collection in mid-April was intended to provide an opportunity to reach sample members in person before many of them became highly mobile over the summer. Many sample members would still be taking classes at a postsecondary institution at that time, and then follow other pursuits when classes ended in May. Experience from the F2 field test indicated that a significant number of sample members became more difficult to locate once spring classes ended. Starting CAPI data collection in April was intended to avoid situations where sample members had to be relocated once they began summer activities, such as working, traveling, or taking summer classes at another institution.

The data collection target for CAPI data collection was to complete 1,500 interviews, or about 9 percent of second follow-up sample members. Because of the flow of data collection across modes, field cases were expected to generally represent those cases that were most difficult to successfully locate and interview. CAPI projections followed a similar pattern as CATI data collection, but with the expectation that it would take some time to reach full production level. Field interviewing typically requires a critical mass of cases to maximize productivity, and the flow of cases from CATI or tracing to field was limited over the first few weeks of CAPI data collection. Field data collection was expected to begin to taper off in July and further still in August, as the number of pending cases ultimately declined. Section 4.3.2.1 presents the results for CAPI interviewing.

CAPI procedures. The approach for CAPI data collection followed the strategy used successfully on B&B:93/2003 and other recent NCES studies. This approach first identified clusters according to the last known ZIP code of the sample members that could potentially be assigned to CAPI interviewing. Then, based on the distribution of cases by cluster, the geographic clusters that had the highest concentration of cases were staffed with one or more field interviewers. In reviewing caseloads across clusters, particular attention was also paid to those with a significant number of sample members identified as dropouts. Both the 2004 and F2 field test data collection indicated that dropouts would generally be more difficult to include in the study, so CAPI data collection attempted to maximize the yield from this subgroup in the sample.

Cases were assigned to CAPI data collection via a cluster control system (CCS). The CCS used geographic information systems based on ZIP codes to map out geographic clusters from where sample members reside. Clusters were identified in stages. Those with the highest concentration of cases were identified quickly and plans to staff them initiated. Initially, the 35 largest active clusters were identified and activated in the CCS. Among other potential clusters,

factors such as the potential caseload, distance from other clusters, and number of dropouts were used to prioritize activating these clusters. These additional clusters were activated as data collection progressed. The cluster identification and activation procedures remained flexible in the early stages of CAPI data collection to ensure that significant numbers of cases, and/or important cases such as dropouts, could be included in field efforts. Furthermore, field clusters were considered in two ranges, either a 50- or 75-mile radius from the center, to determine the optimal configuration of clusters. For example, clusters with high population density were set at the 50-mile radius, while those with a wider distribution of cases were set at the 75-mile radius. This approach was intended to optimize coverage of potential CAPI cases, while at the same time containing the costs associated with collecting data in person.

Across all clusters, the expectation was that a total of about 3,000 cases would eventually be assigned to CAPI data collection. Cases were assigned to CAPI data collection twice weekly on a flow basis, from the start of field data collection through the end of data collection. CAPI cases were composed of cases that met one of three criteria:

1. Unable to locate sample member through CATI and intensive tracing efforts.
2. The sample member declined to participate through CATI efforts or other prior contacts.
3. Substantial efforts were made to contact the sample member via CATI, mail, and e-mail, but had not yet been successful.

When a case was identified as requiring CAPI effort, one of these three codes was assigned to the case so that field supervisors and interviewers knew why each case was assigned to the field—unable to locate, prior refusal, or difficulty in contacting. This procedure assisted CAPI staff in developing an appropriate approach to each case.

After cases identified for CAPI data collection were assigned to specific clusters, the field manager and field supervisors used a web-based Assignment/Transfer System (ATS) to assign cases to specific interviewers. Some areas where sample is concentrated comprised a single cluster with only one field interviewer, while cases in more populous areas were assigned to multiple interviewers in the area. The web ATS also enabled field supervisors to initiate transfer of cases from one field interviewer to another, as necessary.

To assist CAPI interviewers in contacting each sample member, a record of prior CATI and tracing activities was provided for each case assigned to the field. Field interviewers, with the help of their supervisors, develop a strategy for making personal visits to various locations where sample members, parents, and other locator sources are known to have resided at some time. This included visits to addresses obtained during the base-year and first follow-up data collection efforts or from other sources used during previous tracing efforts. If the sample member is known to have attended a high school that did not provide a transcript, the field interviewer could occasionally contact the high school to attempt to confirm or ascertain the sample member's last known address and phone number. Field interviewers also followed up on any leads related to postsecondary institutions the sample member applied to or attended. Further locating steps taken included, for example, searching a postsecondary institution's website and/or contacting the institution by phone to confirm that the sample member was attending the school and to obtain any phone number or address information that was available. If the institution or new address obtained was beyond the interviewer's work area, field supervisors

determined the best course of action for pursuing the case. Next steps included transferring the case back to CATI data collection (when a new phone number was obtained), transferring the case to another field interviewer who was nearer the institution or new address, or sending an interviewer to the sample member's area. Field locators were also added to the CAPI staff beginning in July to assist interviewers with locating the most difficult-to-find sample members.

Field staff carried a Letter of Authorization signed by the NCES project officer to validate their legitimacy as a field interviewer for ELS:2002. The letter was intended to help overcome barriers to participation that are frequently encountered when making "cold" contacts face-to-face, especially with persons who had not previously seen the study materials. Field interviewers also carried copies of the lead letter and brochure so they could quickly provide these study materials to sample members, parents, or others who had questions about the study. In addition, because the introductory statement, respondent confirmation, and informed consent scripts were not built into the CAPI interview, field interviewers carried laminated hard copies of these scripts so that they could be sure to present them appropriately to respondents.

The CAPI interview was conducted on laptop computers via a web-based interface that used personal web server software. A local database resided on each laptop to assist with coding operations and other temporary storage of data during the interview. To maintain consistency across interviewing modes, the CAPI interview was identical to the CATI interview. The same interviewer instructions presented in CATI interviews were included in the CAPI interview. CAPI interviewers also followed the same standardized interviewing procedures as CATI interviewers, including techniques for reading questions, probing insufficient answers, recording answers, and maintaining neutrality with respect to respondents' answers. All CAPI interviewing procedures were conducted within this same set of standardized interviewing techniques established prior to data collection and presented in training.

CAPI interviewers were allowed to administer the interview over the telephone, which produced conditions even more similar to CATI interviewing. While this was not the preferred option, it proved useful for those cases where the sample member could be reached more easily by telephone than in person and was willing to do the interview when contacted. This option was also useful for completing interviews with mobile sample members who temporarily left the field interviewer's cluster area during the data collection period. For sample members who preferred to complete the survey via the Web, field interviewers ensured that they had the necessary information and instructions to complete web self-interviewing. Field interviewers typically allowed these sample members about a week to complete the interview through the Web. If they had not completed the survey during that time, field interviewers then called them back to ask if they had encountered any problems and to offer to complete the interview with them by telephone.

When they found sample members who had not yet participated in the survey, field locators were allowed to assist them to either access the web self-administered interview or call in to complete a CATI interview. If the sample member agreed to complete the interview by either mode, field locators could then wait in the home until the interview was finished to provide the respondent with his or her incentive in person. In these situations, the locator instructed the respondent to alert him as soon as the respondent completed the last item, which was described to him before beginning the interview. Overall, few field interviews were conducted in this manner. In all such cases, the field locator provided the respondent with the incentive as soon as the interview was completed.

At the end of each workday, CAPI interviewers electronically transmitted their completed interview data and status codes for each assigned case to RTI computers. The field transmission ensured that field interviewers regularly delivered data to the main ELS:2002 database in a timely fashion and that interview data could therefore be secured in the main data files. The transmitting system also allowed field interviewers to receive new cases when they were assigned and to capture updates to case information and the survey instrument as necessary.

CAPI staffing and training. Given the challenging nature of CAPI data collection, an experienced and skilled staff of field interviewers was critical to the success of this effort. CAPI staffing began with recruiting and hiring five experienced field supervisors to assist the CAPI manager in managing the field interviewing staff. Recruitment of the field staff targeted veteran field interviewers who had strong past performance on similar studies. In addition, those with extensive experience in locating difficult-to-find respondents and converting initial refusals were favored in staffing CAPI data collection. Because we were not able to exclusively hire interviewers with extensive experience and demonstrated skills, additional interviewer candidates with strong references, computer experience, and strong interpersonal skills were also recruited. In recruiting interviewing staff, we also considered the demographic characteristics of the set of field clusters, so that we could include interviewers experienced in interacting with persons of relevant ethnic and linguistic groups in the cluster, such as native Spanish speakers.

The first field interviewer training was conducted April 9 through April 12 and a second training May 21 through May 24. Both CAPI training sessions were composed of 2 hours of pre-classroom home study, three 8-hour days of training, and a final 6-hour day, for a total of 32 hours. Trainees who were new to interviewing attended an additional 8-hour day of training on general interviewing and computer skills prior to the first day of each training session.

A comprehensive, classroom-based training program was developed for the 2006 CAPI training. Key information on the purpose and goals of ELS:2002 and specific interviewing procedures was compiled in a manual for field interviewers to reference. The training emphasized quality aspects of interviewing (such as instructions on field tracing, enlisting cooperation, and correctly administering the interview) as well as data collection efficiency issues (such as using locating information and prioritizing visits to sample members or parents). Specific training sessions included the following:

- structured practice with the web-based program, CMS, e-mail, and data transmission systems;
- review of case documentation, procedures, and reporting, and administrative requirements;
- standardized interview administration techniques;
- specifics of the laptop computer and the use of the CMS; and
- role-playing exercises to practice administering the interview and gaining cooperation from respondents.

Most of these training modules were conducted by combining a presentation with hands-on practice or group discussions.

In addition to completing the home study exercise prior to training, all field interviewers were required to pass the following certification steps on the final day of training:

- A verbal test on responding to frequently asked questions from sample members and gaining cooperation.
- A mock interview with a trainer as respondent.
- A coding exercise for postsecondary fields of study, postsecondary institutions, and occupations.
- An exercise on selecting appropriate event codes and entering a code in the CMS.

Completing these activities ensured that the CAPI interviewers were well prepared in both administering the interview appropriately and enlisting cooperation in the most challenging situations.

Quality control measures in CAPI. Like CATI efforts, CAPI data collection included multiple procedures to ensure that data quality standards were being maintained. The CAPI task leader and field supervisors closely monitored CAPI production on a daily basis so that they could quickly address production issues and other field data collection challenges. Field supervisors held weekly conference calls with each of their field interviewers to discuss the status of each assigned case and ensure appropriate efforts were being made for each case. During these calls, particular emphasis was placed on handling refusal cases and determining appropriate steps for locating cases. The CAPI manager also held weekly conference calls with each field supervisor to discuss field production and strategies and to communicate any updates on data collection plans.

To maintain control of quality in CAPI data collection, verification interviews were conducted for a sample of each field interviewer's completed interviews. At the end of each CAPI interview, respondents were told that they might be contacted for quality control purposes. Verification calls and interviews were completed by in-house telephone interviewers. Training to conduct verification calls was held on May 8 and verification calls began immediately. Completed CAPI interviews were sampled randomly over the course of data collection and added to a modified CATI-CMS program to schedule the verification calls. The verification interview included a brief set of questions about the procedures followed during the original interview, including the date on which the interview occurred, the mode in which the interview was completed (by telephone or in person), the approximate duration of the interview, and the amount of the incentive paid. In addition, two key factual questions from the 2006 interview were asked again in the verification interview: whether the respondent had held a job for pay since high school, and whether the respondent had attended any postsecondary institutions since high school. Any problems detected through verifications were coded and displayed on a verification report. The report was monitored by the CAPI manager and field supervisors so that issues could be addressed with the field staff member in a timely manner.

4.3.1.4 Intensive Tracing During Data Collection

When sample members were determined to have insufficient contact information to reach them as a result of outbound CATI data collection efforts, more intensive tracing efforts were undertaken. The two primary intensive tracing modes were centralized interactive tracing and field locating. Centralized tracing was conducted by tracing specialists located in RTI's Call Center Services facility. The centralized tracing operations followed a comprehensive and proven set of procedures for locating sample members using Internet searches and telephone

contacts. Tracing specialists were trained exclusively in tracing procedures, resources, and investigative techniques and most had experience in tracing the young adult population that comprised the ELS:2002 sample.

When cases were assigned to CAPI data collection, field tracing was performed by field interviewers or field locators as needed. CAPI staff with experience in locating difficult-to-find sample members were favored in staffing the field effort. Field interviewers and locators relied on well-established techniques to trace sample members in their local communities. Field tracing was also combined with centralized tracing procedures when field locating efforts produced specific information that could be used by tracing specialists to perform online searches.

A third set of tracing procedures added during data collection was peer locating. This involved prompting participants to provide any information they might have about ELS:2002 sample members who attended the same high school. These activities were reserved for the latter half of data collection to assist the data collection team in obtaining leads on sample members who had proven most difficult to locate.

Centralized interactive tracing procedures. A total of 12 tracing specialists and 7 quality control specialists were trained for the interactive tracing effort. All tracing specialists had prior experience conducting interactive tracing, and many had worked on previous rounds of ELS:2002. Tracers and tracing supervisors were trained in a 2-hour session that included background information on sample maintenance activities and prior contacts with ELS:2002 sample members, an overview of interactive tracing procedures, discussion of particular challenges in locating the young adults in the sample, and responding to frequently asked questions when contacting sample members or others. Interactive tracing efforts began on February 27 and continued through September 7, 2006. Throughout data collection, quality circle meetings were held periodically with tracing staff to ensure that best practices were being followed and to discuss strategies for successful tracing.

The starting point for interactive tracing efforts was the contact information provided during the baseline or first follow-up data collection, and any updates to this information acquired through the batch tracing and sample maintenance activities conducted in fall 2005, as described in section 4.3.1.1. All of this information was included in the ELS:2002 locator database. In the BY and F1 rounds, participants had been asked to provide the following information as part of the interview protocol:

- respondent's full name, address, and current telephone number;
- respondent's Social Security number (SSN);
- full name, address, and telephone number of mother/father or female/male guardian of respondent;
- full name, address, and telephone number of a close relative not currently living with the respondent who would always be likely to know how to locate the respondent;
- full name, address, and telephone number of close personal or family friend not currently living with the respondent who would always be likely to know how to locate the respondent;
- respondent's nickname, if any;

- basic occupational information about both parents/guardians; and
- any postsecondary institutions to which they had applied and/or planned to attend.

Although information about postsecondary institutions obtained in the 2004 first follow-up was expected to have limited usefulness, this avenue was explored in tracing efforts. Given the high proportion of sample members expected to be enrolled in postsecondary institutions, any available school directories could be used to help locate sample members.

Another important aspect of tracing sample members was using parent information. Because sample members were now mostly 19 or 20 years old, a significant number had information to facilitate locating them through interactive tracing. For other sample members, parent information was the most useful route to locating sample members.

When CATI interviewing efforts were unable to locate sample members at any of the telephone numbers available, the case was identified as needing tracing. CATI supervisors then reviewed these cases to ensure that all available numbers had been exhausted. This effort varied in complexity, as the number of unique telephone numbers available for each sample member ranged from zero to five (or more).

Once reviewed cases were assigned to interactive tracing, tracing specialists implemented a systematic set of procedures for tracing sample members that had been shown to be effective on prior rounds of ELS:2002 and similar studies:

- Check preloaded information: Check case, source/contact, lead, and case history screens for any relevant information.
- Call all preloaded phone numbers: Verify that all preloaded phone numbers are working and whether the subject can be contacted through these numbers.
- Trans Union: Run "Trace" SSN search if provided. When one does not have SSN, run the "ReTrace" search.
- Experian SSN search/address update: Run SSN search if SSN is provided and address when SSN is not given.
- Fast Data address search: Run address search on the subject and contact for preload and developed contacts.
- Fast Data DA Plus: Run a DA+ search on the generated addresses.
- Fast Data reverse search: Run a reverse search on generated phone numbers not associated with physical addresses.
- Fast Data name search: Run a name search for the subject in the city and state in which there are contacts for the subject. Note: A surname search can also be used to develop information for relatives.
- Accurint: Perform name search, address search, reverse phone search, and neighbor search.
- Repeat any of the above steps as necessary, depending on any leads developed.

These steps were used interactively in that tracing specialists could change the sequence of subsequent steps depending on specific leads found in initial steps. Furthermore, when new

telephone numbers were found through any of these sources, tracers would attempt to confirm that the sample member was in fact residing at that telephone number. Direct confirmation was a key step in the interactive tracing process, as confirming telephone numbers allowed for cases to be returned to CATI data collection immediately without further tracing efforts. When new telephone numbers could not be directly confirmed, tracing specialists would generally conduct additional searches to corroborate this information.

All information obtained for cases through centralized tracing was added to CATI-CMS. The most current locating information for sample members was therefore readily available for additional data collection and tracing efforts. When new telephone numbers were found via interactive tracing (confirmed or unconfirmed), the new numbers would be exported to CATI-CMS for continued CATI data collection efforts. When only address information was obtained for sample members, the information was exported to CATI-CMS, but the case was also identified as requiring CAPI data collection efforts. Cases for which no new information was obtained would also be assigned to CAPI data collection.

Based on experience from the ELS:2002 2004 data collection, the expectation was that as many as 33 percent of sample members would require some level of interactive tracing during data collection. A total of 3,000 cases were assigned to interactive tracing at least once during data collection. Among these, 320 cases were assigned to interactive tracing a second time. Unlocated cases were assigned to a second round of tracing when one of the following conditions was met:

- no field interviewer was currently available in the local area to conduct field locating;
- new information was obtained from field locating that could be used more effectively in interactive tracing; or
- review of the initial tracing effort indicated one or more leads could be pursued further through additional tracing efforts.

Overall, 81 percent of cases traced interactively (one or more times) resulted in new information being obtained. No new information was obtained for 17 percent of cases traced, and work was stopped for the remaining 2 percent of cases. (Stop work orders resulted from cases being finalized prior to the completion of interactive tracing efforts.) Although the overall rate of locating information was relatively high, not all information obtained through interactive tracing was equally useful. For this reason, reports of case status were created for each type of locating information obtained (telephone, address, or both) and whether this information was confirmed. These reports allowed data collection staff to more accurately assess the ultimate value of information obtained via interactive tracing for locating and interviewing sample members. The overall response rate was highest for those cases where both a new address and telephone number were obtained through interactive tracing (83 percent), and lowest for those cases where only an unconfirmed address was found (65 percent). For cases where only sample members' parents could be located through interactive tracing, the success rate was even lower (61 percent).

Field tracing procedures. Despite the general success of finding new contact information for sample members via centralized interactive tracing, useful information could not be obtained for a significant portion of second follow-up sample members. For these cases, the next step was

to assign them to CAPI interviewers for field locating efforts. Locating issues and strategies were included as part of CAPI training.

The starting point for field locating was the last known address for each sample member. For each case assigned to CAPI data collection, a record of interactive tracing results as well as the CATI call history was provided to the interviewer. CAPI interviewers, with the assistance of their supervisors, developed strategies for visiting various locations where sample members, their parents, and other locator sources are known to have resided at some point. Initial field tracing steps included revisiting or recontacting leads documented by centralized tracing efforts, calling sample members' and parents' old phone numbers, and visiting sample members' and parents' previous addresses. One advantage of field locating over centralized tracing was that field staff could use information obtained locally in combination with the information provided from previous tracing and data collection efforts to determine where sample members most likely resided. Beginning in late July, 13 experienced field locators were hired and trained to augment field locating efforts.

When field interviewers or locators made contacts as part of field locating efforts, they asked a set of questions about the sample member's spouse or partner, work situation, recent moves, and related questions to generate further leads. Standard questions included:

- Is the sample member married? What is the spouse's name?
- Does the sample member work? Where? What kind of work does he or she do?
- Does the spouse work? Where? What kind of work does he or she do?
- When did the sample member move? Do you know where?
- Did the sample member own or rent the home?
- Does the sample member attend a local church?

If these questions did not produce useful information or leads, field staff would also ask contacts for suggestions on the most likely way to contact the sample member or parents. Field interviewers used a checklist of potential sources to document the steps taken to locate sample members, covering a variety of possible contacts and local institutions that could be useful for obtaining contact information or other leads. These records were useful for documenting efforts for additional field locating or centralized tracing steps.

In addition, field staff frequently attempted to locate sample members by using local information or leads obtained for conducting Internet searches. Websites such as http://www.whitepages.com/, http://zabasearch.com/, and http://theultimates.com/ often produced good results, especially for looking for relatives of parents and other family members. Using information obtained in sample members' communities, such as information from neighbors, often provided an advantage for field staff in searching for sample members.

Peer locating procedures. One further tracing activity implemented during data collection was peer locating. Because of the challenges of locating highly mobile ELS:2002 sample members, peer locating was initiated in May. Peer locating involved two related activities: sending e-mail messages to participants asking for their assistance in locating sample members who attended the same high school and conducting outbound prompting calls to selected participants in order to request their assistance with locating these classmates. For both

participants and interviewers who were conducting prompting calls, a special, secure web interface was established to enter any contact information for pending sample members. This web application was accessible only to respondents who were contacted as part of the peer locating effort.

The first step in peer locating involved sending e-mails to 9,300 participants on May 19, 2006. Those receiving e-mails included all respondents who attended the same high school as at least one second follow-up sample member who had not yet completed the interview. The e-mail message explained to participants that the data collection team was seeking their assistance in locating pending sample members and provided instructions on how they could provide this information, including a direct link to the website. When respondents logged on, they were prompted with a list of one or more classmates who had not yet participated in the survey. Once these sample members did complete the survey, they were automatically removed from the peer locating list.

After allowing participants approximately 4 weeks to respond to the request for information on their peers, prompting calls were initiated to those respondents who had not yet logged on to the website to attempt to provide information for their listed classmates. Prompting calls began in late June, and included 7,500 participants. Up to three prompting calls were made to each respondent. Once successful contact was made with participants, prompting calls were completed. Otherwise, three attempts were made for each respondent.

Peer locating efforts targeted a total of 3,600 ELS:2002 second follow-up sample members who had not yet completed the interview and had at least one classmate who had completed the interview prior to May 19. A total of 1,600 unique pieces of information for these pending cases was obtained from peers. This information pertained to 1,000 sample members, or about 28 percent of the cases targeted.

Data collection staff then evaluated the information received through peer locating efforts in terms of its usefulness. Overall, about 20 percent of peer locating information was judged as likely to be useful, which included new telephone numbers, updated addresses, or other contact information. Another 70 percent of the information was evaluated as potentially useful, such as information on which postsecondary school the sample members may have attended. The remaining 10 percent of the information was considered not useful because it did not indicate where the sample member was currently living or how the person could be contacted. Information judged likely or potentially useful was then added to the locator database and/or communicated to interviewing staff.

4.3.1.5 Data Collection Procedures for Initial Refusals and Difficult Cases

Another important challenge in planning data collection was developing procedures for sample members who initially refuse to participate or who otherwise prove difficult to include in the study. As indicated in section 4.3.1.2, the design of the incentive plan considered factors likely to increase the difficulty of including certain sample members, such as those who did not participate in the previous wave and those who had dropped out of high school. The incentive plan was intended to reduce the potential for sample members to hesitate or refuse to participate when first contacted about the data collection. Because the incentive plan could not avert initial hesitation or refusal among all sample members nor address all reasons for hesitation or refusal, procedures were needed to overcome hesitation and avoid refusals among sample members. In addition, because not all refusal or difficult situations could be avoided, contingency procedures

were needed to address initial refusals or other difficult-to-complete cases during data collection. This section describes the procedures in place to avoid refusals, manage initial refusals, and handle other difficult situations.

Procedures for avoiding refusals. Procedures for avoiding refusal situations included three primary sets of activities: interviewer training sessions, web/CATI quality circle meetings, and sample management. Efforts to avoid sample member refusals began in the training sessions of web help desk and telephone and field interviewing staff. Training modules addressed common reasons for reluctance or refusal, strategies to address potential refusal situations, and consideration of specific reluctance or refusal statements and behaviors. Presentations, discussion, role-playing exercises, and a team competition were all used to prepare interviewers to address potential refusal situations. These training modules included specific objections from sample members or parents and potential interviewer responses that were directly based on experiences from prior rounds of ELS:2002 and other current education studies of young adults.

Another important focus of interviewer training was addressing potential reluctance or specific objections among gatekeepers. Experience from prior rounds of ELS:2002 demonstrated the ways in which parents and other household members can either help or hinder efforts to contact sample members. Training modules also focused discussion and exercises on how interviewers can successfully address common gatekeeper concerns and objections. To assist CATI interviewers, the CATI-CMS program included scripted probes for interviewers to use when asking parents' or other contacts' cooperation in reaching sample members. Because many sample members had completed high school and moved out of their parents' household, gaining parent assistance in contacting sample members was often the first step in the survey participation process.

These training sessions all followed the same general strategy for addressing reluctant sample members or gatekeepers, including the following:

- understanding the reason(s) for the subject's or gatekeeper's reluctance as quickly as possible;
- being prepared to address the concern(s) quickly and directly;
- focusing responses on why the sample member's participation is important to ELS:2002; and
- using an effective tone and maintaining a professional approach.

This strategy was illustrated through specific examples used in training modules.

In addition, all interviewers were required to complete an exercise in responding to sample member or gatekeeper concerns as part of the certification process mentioned in section 4.3.1.3. This certification process reinforced using the refusal avoidance strategy to communicate the importance of sample members' participation in ELS:2002. The most important points interviewers were trained to communicate to sample members and other contacts included:

- reminding sample members and parents of their previous participation;
- the importance of sample members' continued participation;
- the importance of ELS:2002 for education in the United States;

- the incentive payment provided to sample members;
- explanations of the 2006 data collection procedures and options; and
- a toll-free number to talk with the data collection manager about the study.

Key talking points and refusal avoidance strategies were regularly reinforced in quality circle meetings held with help desk and telephone interviewing staff. As mentioned in section 4.3.1.3, QC meetings were held biweekly during 2006 data collection to ensure procedures were being followed correctly. QC meetings provided a forum for interviewers to discuss specific examples of reluctance or refusal responses among sample members and possible steps to address these concerns. After the first few QC meetings focused on basic issues, data collection managers began to regularly add a session at the end of each QC meeting devoted to role-playing potential refusal situations. These sessions provided regular practice and discussion for interviewing staff so that they were prepared to address these situations effectively. For CAPI data collection, weekly calls between field supervisors and field interviewers addressed the same kinds of issues in one-to-one conversations.

Sample management activities in CATI data collection were also an important part of refusal avoidance procedures. Call scheduling procedures were designed to avoid inundating households in the sample with too-frequent calls. For example, when answering machines were reached and a message left, the CATI-CMS call scheduling system held these cases for at least 3 days to give sample members or their parents some time to return the call. When sample members or parents requested a callback for a specific day and time, interviewers entered these appointments in CATI-CMS so that the appointment cases would be delivered to interviewers at the appropriate time. Telephone supervisors were continually aware of the need to keep all appointments, and monitored the status of upcoming appointments to ensure all appointments were covered. These procedures ensured that appointments were kept regularly, which was a significant issue for sample members and parents with busy schedules. Another important feature of CATI-CMS that provided assistance in avoiding refusal situations was the call history log. After each call, interviewers entered relevant information about the results of the call and any interaction with sample members or other contacts in this log. These notes ensured that interviewers who made subsequent calls to contact sample members were aware of the results of previous calls. The call history log allowed interviewing staff to be sensitive to any concerns sample members, parents, or other contacts had about the 2006 data collection process and to be prepared to address those concerns in subsequent contacts.

CAPI staff also used sample management techniques to avoid refusal situations. A key difference between CAPI and CATI was that field efforts to avert refusals were based on collaborations between field supervisors and field interviewers. Field interviewers worked with their supervisors to develop an individual approach to each case based on the reason the case was sent to the field, the CATI call history for the case, and the results of any initial contact attempts by the CAPI interviewer. Field staff maintained detailed documentation of contact attempts on a "Record of Actions" form and also entered this information in the field CMS. Field interviewers could then review this information when planning future contacts with each sample member and, if the case was transferred to another interviewer, provide information for those subsequent contact attempts.

Procedures for converting refusals. Despite the procedures in place to avoid refusal situations, a total of 660 initial refusals occurred among sample members. Another 1,400 initial

refusals occurred across all contact attempts, including refusals where parents or other contacts attempted to decline on the part of sample members. Whenever a call resulted in a refusal, CATI interviewers followed a predetermined set of steps to classify the refusal situation. CATI-CMS produced a series of screens that allowed interviewers to specify the following information:

- person who refused (sample member or other);
- point at which the refusal occurred (prior, during, or after the introduction);
- strength of refusal (mild, firm, or hostile); and
- any specific reasons mentioned for the refusal.

Most of the initial refusals by sample members (70 percent) were coded as "mild" by interviewers, indicating that in most cases interviewers simply contacted the sample member at a time or in a situation where he or she was not prepared to participate. Table 41 provides a breakdown of the point of refusal and primary reasons for sample member refusals, when provided. This table records both the timing and reason(s) for refusals, so cases are often represented two or more times. The first three rows provide support to the conclusion that a significant number of initial refusals were often quick interactions where sample members, parents, or other contacts either did not fully understand the purpose of the call or were not in a situation where they were able to participate. Just about half of all initial hangups occurred prior to the reading of the introductory script, during the introduction, or just after the introduction when the sample member's identity was being confirmed. Among reasons for refusal, the two most common reasons reported were not being interested in general (22 percent) and specifically not being interested in participating in ELS:2002 again (21 percent). Another 12 percent of sample members indicated that they were too busy to participate and 14 percent provided various other reasons for declining the interview. For some 10 percent, no information was provided as to either the timing of or reason for refusal. These results provided an overview of the nature of refusals that data collection managers and interviewing staff used to adapt procedures for converting refusals.

Table 41. Timing of and reasons for initial sample member refusals: 2006

Refusal outcomes	Number	Percent
Hung up before introduction	50	6.7
Hung up during introduction	200	35.6
Hung up during sample member verification	50	7.6
Too busy/no time	80	12.0
Not interested (no mention of ELS:2002 study)	150	21.8
Not interested in participating in ELS:2002 again	100	20.9
Concerned about purpose of study	#	0.7
Concerned about how long survey will take	#	1.2
Concerned about how their contact information was obtained	#	0.6
Other reason specified	100	14.1
No information reported	70	10.5

Rounds to zero.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Interviewing staff used both general strategies and the specific information in CATI-CMS, including the call history log, to develop a refusal conversion approach to each individual case. After a call resulted in a refusal and the information about the interaction was entered,

CATI-CMS moved the case to a special refusal queue. Cases in the refusal queue were held for at least 1 week after the initial refusal before being made available by the call scheduler for subsequent refusal conversion attempts. In practice, the time interval between the initial refusal and the next contact was often longer than 1 week, as multiple subsequent calls were often required to contact these sample members again.

Delaying subsequent contact attempts was a key sample management procedure used to maximize the success of refusal conversion efforts. This break provided a short period to allow sample members to reconsider their participation in the study and, in some cases, to be in a more favorable situation to participate. For the same reason, CAPI interviewers who were assigned initial refusal cases from CATI data collection typically waited a week before attempting to contact these sample members. All refusal cases assigned to field data collection were reviewed carefully by the field supervisor and field interviewer. Field staff would review the statements made by the prospective respondent or gatekeeper when they declined participation and develop a refusal conversion approach for each individual case. The approach of field staff sometimes included having the supervisor contact the household first, in some cases, or transferring the case to another CAPI interviewer when the original interviewer was unable to make progress with the case.

Because data collection staff anticipated that a significant number of refusals would ultimately transpire, plans were made early in data collection to conduct specialized refusal conversion training sessions for telephone interviewing staff within a few weeks after the start of outbound CATI calling. The first refusal conversion training was conducted about 3 weeks after outbound CATI data collection, and a second training session was held 2 weeks later. For both training sessions, data collection staff selected interviewers with strong performance ratings to attend these trainings. Interviewers were also identified based on qualitative feedback from telephone supervisors and monitors. As a general rule, interviewers selected to be refusal conversion specialists were interviewers who had demonstrated skills in enlisting cooperation among sample members and avoiding initial refusals. The training sessions emphasized specific refusal conversion techniques tailored to the ELS:2002 sample of young adults, including overcoming objections, addressing concerns of gatekeepers, and providing alternatives for participation. Both group discussions and individual role-playing exercises were used in refusal conversion training. Only interviewers who had successfully completed one of these training sessions were allowed to call initial refusals. Section 4.3.2.3 presents the results of refusal conversion efforts across all modes of data collection.

Procedures for addressing other difficult situations. In addition to sample member refusals, 2006 data collection efforts encountered other difficult cases. As described in section 4.3.1.2, a number of criteria (including refusal) were used during data collection to designate cases as difficult. The most common nonrefusal situations that led to the difficult case designation included the following:

- more than 20 contact attempts were made without completing the interview;
- the case was submitted to intensive tracing because the sample member could not be located;
- the case was assigned to field data collection, either because the sample member could not be located or could not be contacted by telephone; or
- the sample member had not completed the interview as of June 15, 2006.

As noted in section 4.3.1.2, meeting any of these criteria resulted in an increase in the incentive amount. Section 4.3.1.4 describes procedures for intensive field tracing of sample members who were difficult to locate. This section describes procedures for addressing other kinds of difficult situations, especially sample members who were difficult to reach by telephone.

Similarly to refusal conversion efforts, a number of additional procedures beyond the incentive increase were implemented to manage difficult-to-reach sample members. When cases were designated as difficult, CATI-CMS moved them to a special queue. Cases in the difficult queue were assigned to interviewers in a similar manner as refusal cases. Only telephone interviewers who had demonstrated skills in enlisting cooperation among sample members and handling difficult situations were assigned to call cases in the difficult queue. As with refusals, interviewing staff used both general strategies and the specific information in CATI-CMS, including the call history log, to develop an approach to address each difficult case. Based on the criteria established by the data collection staff, all pending nonrefusal cases were moved to the difficult queue on June 15, 2006. At that point, all CATI interviewing staff were prepared to call difficult cases. Like refusal situations, difficult nonrefusal scenarios were discussed with telephone interviewers in QC meetings and with field interviewers in regular calls with field supervisors.

One of the most common reasons for nonrefusal cases to become designated as difficult was inability to contact sample members at any of the telephone numbers available in the locating database. An important challenge in CATI data collection efforts was overcoming call-screening behavior. A significant number of households did not respond to telephone calls even after multiple attempts had been made and answering machine messages had been left. This challenge was exacerbated by the fact that contact information for many cases was initially limited only to phone numbers for sample members' parents and other relatives. As a result, telephone interviewers had to make contact with the parent households first to determine a current number where the sample member could be reached. For this reason, the number of calls required to contact sample members by phone was often increased, particularly when the parent households were screening calls and not responding to answering machine messages. A third factor that increased the difficulty in reaching sample members was that many were only reachable by cellular phone. Even when parents or other contacts provided cell phone numbers for sample members, many sample members were concerned about the costs of using their cell phone to complete a CATI interview. This same concern about costs also led some parents to be reluctant to provide cell phone numbers for sample members to telephone interviewing staff. All of these factors combined to increase the challenges of contacting some sample members and completing interviews by telephone, resulting in a significant number of cases being designated as difficult.

Similarly to circumstances for refusal cases, the call history log in CATI-CMS was an important resource for telephone interviewers in attempting to contact difficult cases. In addition to detailed notes, the call history log provided CATI staff with the distribution of call attempts across all numbers and the results of each call attempt. Interviewing staff could then use this information to determine the telephone number where contact was most likely to occur and the day and time when contact was most likely to occur. Likewise, the call history indicated numbers where productive contact had and had not been made, so that interviewers could prioritize calling across multiple telephone numbers. As calling attempts to reach difficult cases continued to

prove unproductive, data collection managers increasingly assigned such cases to CAPI data collection.

Field staff experienced significant success with making telephone contact with difficult cases that had not been successful in CATI. At least part of this success likely resulted from attempts to contact households that had previously been screening calls from telephone interviewers. All calls from RTI's Call Center provide the same telephone number in caller ID systems. When field interviewers called these same households, a new, local number would appear in caller ID systems. The novelty of a new phone number, voice, and/or approach likely contributed to field interviewers' success in contacting difficult cases after substantial CATI data collection efforts had been unsuccessful. Consistent with the data collection plan, switching the most challenging cases from CATI to CAPI modes was often an effective strategy for contacting difficult-to-reach sample members. When telephone contacts did not initially prove successful for field interviewers, a personal visit to the sample member or his or her parents' homes was the next step. Field interviewers also used in-person contact as the first step to reach some difficult cases, especially when a high number of prior calls had proven unsuccessful and no alternative telephone numbers were available. Personal visits not only increased the likelihood of face-to-face contact with sample members or their parents, but also proved effective for obtaining updated telephone numbers from parents or other contacts.

4.3.2 Data Collection Results: Outcomes and Indices of Data Quality

The following section provides select data collection and data quality results. Several data collection outcomes are discussed, including:

- response rates by various subgroups;
- refusal and conversion rates;
- distribution of respondents by month of interview;
- distribution of respondents by questionnaire administration mode;
- interview completions by incentive type;
- telephone interviewer hours expended and call counts;
- field interviewing results;
- interview completion time;
- analysis of field of study and occupation recoding; and
- interviewer error rates.

4.3.2.1 Outcomes: Case Response Rates by Subgroup and Data Collection Mode

Response rates by subgroup and mode of administration are presented in this section. For the second follow-up (but not the base year or first follow-up), the response rate is a conditional

one, based on the cases that were fielded.[54] In addition, refusal and refusal conversion rates are reported for both the sample member and the gatekeeper. Interview completions by select subgroups (such as sex, race/ethnicity, socioeconomic status (SES) quarter, first follow-up response status, and respondents classified as ever having dropped out) are provided overall and by mode.

Overall response rates. The ELS:2002 second follow-up sample consisted of 16,400 members overall. The sample represents a subset of the combined population of 10th-graders in the spring term of 2002 and 12th-graders in the spring term of 2004. Some members belong only to the 10th-grade population, some only to the 12th-grade population, but most belong to both. Of the total sample, approximately 15,900 (97 percent) were considered to be in-scope for the 2006 round. Cases classified as permanently out of scope (deceased, sampling errors) or temporarily out of scope (unavailable for duration of study, out of the country, incapable, incarcerated, institutionalized) were not counted in the response rate.[55]

Second follow-up response rates by select characteristics are presented in table 42. Weighted and unweighted completion rates[56] are provided for demographic subgroups in addition to various student and school characteristics associated with the base-year and first follow-up rounds. Response rates for each subgroup are based on the number of eligible sample members who completed the interview. Completed cases included about 14,200 fully and partially completed web and interviewer-administered interviews. Weighted response rates were calculated using the design weight (i.e., the base weight—the weight that reflects the selection probability but has not been adjusted for nonresponse and indeed is available for respondents and nonrespondents alike). The weighted response rate, therefore, represents the proportion of the combined 10th- and 12th-grade population that was in-scope for the second follow-up, was fielded, and that responded.

[54] An unconditional response rate would include cases that were not fielded in the second follow-up: double (base-year + first follow-up) nonrespondents, senior freshening sample nonrespondents, and sample members who withdrew from the study. The response rate as reported here excludes these unfielded cases, that is, it is conditional on the fielding of the case. The unconditional weighted response rate was 84.5 percent overall. The weighted conditional response rate (response rate as used in second follow-up reporting in this document) was 88.4 percent. Ineligible (permanently or temporarily out-of-scope cases) count neither in the case completion rate nor the response rate calculation though their numbers have been documented.

[55] In addition, a handful of previously cooperating sample members asked to be removed from the sample.

[56] Weighted response rates using the base weight are presented because of the importance of population estimation and because NCES survey response standards are based on weighted completions. On the other hand, this chapter's *methodological* tables show unweighted proportions, because of their different focus.

Table 42. Response rates, by select characteristics: 2006

Subgroup	Number eligible	Number of respondents	Unweighted percent	Weighted percent
Total	15,900	14,200	89.1	88.4
Sex				
Male	7,800	6,800	87.2	86.2
Female	8,100	7,300	90.9	90.5
Race/ethnicity[1]				
American Indian or Alaska Native	130	100	87.2	87.5
Asian or Pacific Islander	1,600	1,400	87.4	87.2
Black or African American	2,100	1,900	87.9	87.4
Hispanic or Latino	2,400	2,100	86.4	85.7
More than one race	750	670	89.3	88.0
White	8,900	8,000	90.4	89.5
Socioeconomic status (SES)				
Lowest quarter	3,800	3,300	87.0	86.4
Second quarter	3,800	3,300	87.8	86.5
Third quarter	3,900	3,400	89.0	88.4
Highest quarter	4,500	4,100	92.1	92.5
F1 response status				
F1 respondents	14,700	13,300	91.0	90.2
F1 nonrespondents	1,200	830	66.5	67.4
"Ever dropped out" as of F1[2]	1,200	1,000	85.2	85.0
Student characteristics				
Movers[3]	1,700	1,400	82.5	81.6
Stayers[4]	11,900	10,800	91.2	90.8
Early graduates[5]	660	580	87.6	85.6
Dropouts[6]	830	700	82.8	83.1
Sophomore cohort	15,700	14,000	89.1	88.4
Senior cohort[7]	13,100	12,000	91.3	90.6
BY school sector				
Public	12,500	11,100	88.6	88.2
Catholic	2,000	1,800	92.3	92.4
Other private	1,400	1,300	89.2	88.5
BY school region				
Northeast	2,900	2,600	88.9	89.4
Midwest	4,000	3,600	90.6	89.6
South	5,800	5,200	89.2	88.8
West	3,200	2,800	87.3	85.7

See notes at end of table.

Table 42. Response rates, by select characteristics: 2006—Continued

Subgroup	Number eligible	Number of respondents	Unweighted percent	Weighted percent
BY school locale				
Urban	5,400	4,800	88.8	87.2
Suburban	7,600	6,800	89.1	88.8
Rural	2,900	2,600	89.6	89.1

[1]All race categories exclude individuals of Hispanic or Latino ethnic origin.

[2] Classified as "ever dropped out" as of first follow-up (F1) if at least one of the following conditions was met: school reported that respondent had dropped out of school at any one of the enrollment status updates, respondent was a dropout as of spring term of 2004, or respondent was an alternative completer, that is, earned a GED on or before March 15, 2004.

[3] Includes transfer and homeschooled students. Classification groups reflecting enrollment status (movers, stayers, early graduates and dropouts) were created using a combination of the variable F1QSTAT (for first follow-up respondents), and F1ENRFIN (for first follow-up nonrespondents—spring-term 2004 enrollment status was generally known for nonrespondents, but when unknown was imputed).

[4] Includes students still attending base-year school in spring term of 2004.

[5] Received diploma, GED, or certificate of attendance on or before March 15, 2004.

[6] Completed (respondent) or would have completed (nonrespondent) F1 dropout questionnaire.

[7] Includes spring-term 2004 freshened seniors and sophomore cohort members who remained in modal grade sequence (12th grade) 2 years later.

NOTE: Detail may not sum to totals because of rounding. BY = Base year. GED = General Educational Development credential. Response rate calculation excludes those cases that are permanently out of scope (deceased) or temporarily out of scope (incapable, or unavailable for duration of second follow-up data collection: e.g., out of the country, incarcerated or institutionalized). Total number of permanently or temporarily out-of-scope second follow-up sample members = 460. In addition, unfielded cases are not counted in the denominator of the response rate for the second follow-up. In addition to a handful of sample members who asked to be withdrawn from the study, the following in-scope sample members were not fielded in 2006: double (base-year + first follow-up) nonrespondents (n = 330) and first follow-up freshened senior nonrespondents (n = 40).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Of the approximately 15,900 eligible sample members, about 14,200 completed the ELS:2002 second follow-up survey for an overall unweighted response rate of 89 percent. The overall weighted response rate was 88 percent. Eighty-six percent of males and 91 percent of females completed the interview (weighted). Response rates across racial/ethnic subgroups ranged from a weighted 86 to 91 percent, with White respondents at the high end. Response rates by SES quarter ranged from 86 to 93 percent (weighted), with highest SES quarter respondents at the high end.

The greatest variability in response rates is in the first follow-up response status. As expected, and due to difficulty in locating and contacting (as well as to their presumably higher nonresponse propensities) a large number of first follow-up nonrespondents did not participate in the second follow-up study. Of eligible first follow-up nonrespondents, 67 percent (weighted) completed the interview. However, the weighted response rate for those who had responded in the first follow-up was 90 percent.

Dropouts also historically have been a challenging group to survey. Maintaining the representativeness of this small, select subgroup is critical because the policy relevance of dropouts is high. Two response rates are provided for different classes of dropouts: those identified as ever having dropped out (who were offered a higher incentive) in the second follow-up and a subset of this group—those who completed (or were eligible to complete) the

dropout questionnaire in the first follow-up.[57] Of those offered the "ever dropped out" incentive, the weighted response rate was 85 percent. For first follow-up dropouts, the second follow-up weighted response rate was 83 percent.

Response rates for additional *respondent types* (as determined by completed questionnaire type for first follow-up respondents or assumed questionnaire type for first follow-up nonrespondents) and *cohort type* (2002 sophomore cohort or 2004 senior cohort) are also provided. Weighted response rates by the four first follow-up respondent types ranged from 82 to 91 percent, with "stayers" (those who, in 2004, remained at the 2002 base-year school) at the high end and "movers" (those who transferred to a new school) at the low end. Weighted cohort response rates were similar: 88 percent of those belonging to the 2002 sophomore cohort and 91 percent of those belonging to the 2004 senior cohort participated in the 2006 data collection.

Base-year school characteristics (sector, region, and locale) were also used to classify sample members. Weighted response rates by base-year school sector ranged from 88 to 92 percent, with respondents from Catholic schools at the high end. Weighted regional response rates ranged from 86 to 90 percent, with respondents from the Midwest at the high end. Weighted response rates by school locale ranged from 87 to 89 percent, with respondents from urban schools at the low end.

Refusal and conversion rates. Sample members may refuse to participate for a variety of reasons, including being too busy, not being interested, or having a misconception of what is involved. In addition to refusals made directly by the sample member, gatekeepers—such as a parent or spouse—may refuse to provide access to the sample member or to share locating information. Table 43 and table 44 present unweighted[58] refusal and conversion rates for the 2006 data collection. Table 43 includes both sample member and gatekeeper refusals. Table 44 includes only sample member refusals. A comparison of the refusal rates illustrates the extent to which gatekeepers affect response rates.

[57] To be classified as a dropout in the ELS:2002 first follow-up (F1), one had to be a sophomore cohort member who had been out of school at the time of the F1 data collection for at least 4 consecutive weeks not due to accident or illness, or a returnee who had been in school less than 2 weeks after a dropout episode of 4 consecutive weeks or more. The class of those with "ever dropped out" status is broader in that it also includes students identified by school personnel as out-of-school in tracing who had returned to school by the spring term of the 2003–04 school year and were therefore not classified as sophomore cohort dropouts eligible for the dropout questionnaire but rather as students. An additional group included in the "ever dropped out" category comprises students who had left school and earned a GED prior to March 15, 2004, but had not earned a high school diploma.

[58] Readers are reminded that while both weighted and unweighted percentages were calculated for the *completion rate tables* (because of the importance of weighted data to population estimation) the *methodological tables* (which are concerned not with national estimates but rather with the characteristics and behavior of survey respondents) display unweighted percentages only.

Table 43. Sample member and gatekeeper refusal and conversion rates, by prior response status: 2006

Prior response status	Total	Percentage of sample ever refused F2 interview	Percentage of sample interviewed, after refusal
Total	15,900	12.8	7.8
F1 respondents	14,700	12.4	7.9
F1 nonrespondents	1,200	18.4	6.6

NOTE: Detail may not sum to totals because of rounding. F1 = first follow-up; F2 = second follow-up. Percentages are unweighted. Second follow-up response rate calculation excludes those cases that are permanently out of scope (deceased) or temporarily out of scope (incapable, unavailable for duration of second follow-up data collection: e.g., out of the country, incarcerated or institutionalized). Total number of permanently or temporarily out-of-scope second follow-up sample members = 460. Also, unfielded cases are not counted against the response rate. In addition to a handful of sample members who asked to be withdrawn from the study, the following in-scope sample members were not fielded in 2006: double (base-year + first follow-up) nonrespondents (n = 330) and first follow-up freshened senior nonrespondents (n = 40).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Table 44. Sample member only refusal and conversion rates, by prior response status: 2006

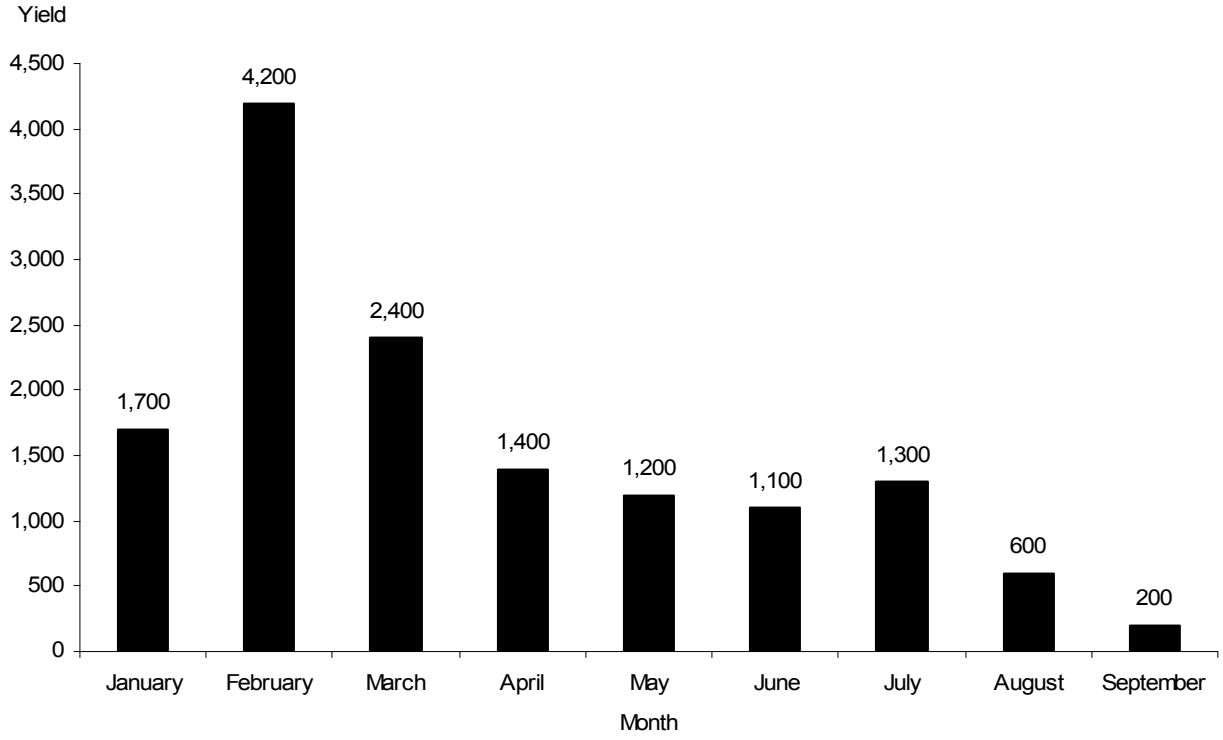
Prior response status	Total	Percentage of sample ever refused F2 interview	Percentage of sample interviewed, after refusal
Total	15,900	8.2	4.3
F1 respondents	14,700	7.8	4.4
F1 nonrespondents	1,200	13.0	3.9

NOTE: Detail may not sum to totals because of rounding. F1 = first follow-up; F2 = second follow-up. Percentages are unweighted. Response rate calculation excludes those cases that are permanently out of scope (deceased) or temporarily out of scope (incapable, unavailable for duration of second follow-up data collection: e.g., out of the country, incarcerated or institutionalized). Total number of permanently or temporarily out-of-scope second follow-up sample members = 460. Also, unfielded cases are not counted against the second follow-up response rate. In addition to a handful of sample members who asked to be withdrawn from the study, the following in-scope sample members were not fielded in 2006: double (base-year + first follow-up) nonrespondents (n = 330) and first follow-up freshened senior nonrespondents (n = 40).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Completions by date (month) of administration. The ELS:2002 field period was relatively lengthy, beginning in January 2006, and ending in September 2006. The point in time at which a respondent was interviewed may affect the data collected—for example, a change in enrollment status as of April might be recorded for a sample member interviewed in June, but not for a sample member interviewed in March. Figure 4 shows the distribution of respondents by month of interview.

Figure 4. Distribution of respondents by month of interview: 2006



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Completions by mode of administration. The ELS:2002 second follow-up survey was multimodal. Three modes of administration were used: self-administered web, and interviewer administration via CATI and CAPI.

Two caveats concerning mode analyses should be entered at the outset. First, no analysis of mode of administration effects on individual survey items was conducted. This is because the validity of such an analysis would depend on random assignment of respondents to modes, and this was not a practical methodology for the survey. Second, while to a great extent mode was "self-assigned"—that is, sample members had the option of selecting web self-administration, or refusing it in preference to CATI or CAPI—not everyone had equal opportunity to do so. For example, sample members who were initially unlocatable and had to be traced had less opportunity to complete a web interview: calendar time had elapsed, early completer incentives for web self-administration had normally expired, and typically the hard-to-locate sample members were urged to complete a telephone or in-person interview at first contact. Certain respondent types—for example, first follow-up nonrespondents and dropouts—were more likely to be hard to locate, and certain demographic subgroups associated with these statuses thus had less opportunity to opt for web self-administration. That said, given the magnitude of differences, there remains evidence that web self-administration was more attractive to some groups than to others.

Table 45 provides the unweighted distribution of completed interviews by mode of administration. Some 47 percent of completions were achieved via self-administered web questionnaire. Some 43 percent were conducted in CATI, and 10 percent were gathered via

CAPI. When combining the CATI and CAPI modes, a little more than half of the cases were interviewer-administered and a little less than half self-administered (53 percent versus 47 percent).

Table 45. Distribution of respondents, by select characteristics and mode: 2006

Subgroup	Total	Web		CATI		CAPI	
		Number	Percent	Number	Percent	Number	Percent
Total	14,200	6,700	47.4	6,100	43.0	1,400	9.5
Sex							
Male	6,800	3,000	43.5	3,200	46.1	720	10.5
Female	7,300	3,700	51.2	2,900	40.2	630	8.6
Race/ethnicity[1]							
American Indian or Alaska Native	120	40	32.8	60	55.2	10	12.1
Asian or Pacific Islander	1,400	800	56.5	500	36.5	100	7.0
Black or African American	1,900	490	26.3	1,100	58.1	290	15.6
Hispanic or Latino	2,100	680	33.4	1,000	50.0	340	16.7
More than one race	670	300	45.0	300	44.0	70	11.0
White	8,000	4,400	54.8	3,100	38.6	530	6.6
Socioeconomic status (SES)							
Lowest quarter	3,300	1,000	31.6	1,800	53.4	500	15.1
Second quarter	3,300	1,400	42.2	1,500	45.5	410	12.3
Third quarter	3,400	1,700	50.7	1,400	41.2	280	8.1
Highest quarter	4,100	2,500	61.6	1,400	34.2	170	4.1
F1 response status							
F1 respondents	13,300	6,500	48.5	5,700	42.9	1,100	8.6
F1 nonrespondents	830	250	30.3	370	44.4	210	25.3
"Ever dropped out" as of F1[2]	1,000	240	24.7	550	56.3	180	18.6

[1] All race categories exclude individuals of Hispanic or Latino ethnic origin.

[2] For "ever dropped out," classified as dropout if at least one of the following conditions was met: school reported that respondent had dropped out of school at any one of the enrollment status updates, respondent was a dropout as of spring term of 2004, or respondent was an alternative completer; that is, earned a GED on or before March 15, 2004.

NOTE: Detail may not sum to totals because of rounding. Provided percentages are unweighted and based on total number of respondents within row. F1 = first follow-up. CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview. GED = General Educational Development credential.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Across all subgroups, completions ranged from 25 to 57 percent for self-administered web, 34 to 58 percent for CATI, and 4 to 25 percent for CAPI. Many subgroups showed different mode of administration propensities, and several subgroups significantly differed from each other in this respect. Mode differences by sex showed that more females completed self-administered web questionnaires than did males (51 percent versus 43 percent [$z = 9.17$, $p < .01$]).

Several racial/ethnic subgroups had higher proportions of completions via CATI/CAPI than self-administered web, including American Indian (67 percent versus 33 percent, $z = 3.71$, $p < .01$); Black (74 percent versus 26 percent, $z = 20.49$, $p < .01$); Hispanic (67 percent versus 33 percent, $z = 15.06$, $p < .01$); and more than one race (55 percent versus 45 percent, $z = 2.58$, $p < .01$). Asian and White respondents, however, were more likely to complete the self-administered web instrument—for Asians, 57 percent of completions were web self-administrations, as contrasted to 43 percent as interviewer administrations ($z = 4.90$, $p < .01$). For Whites, 55 percent were web completions, and 45 percent ($z = 8.57$, $p < .01$) CATI or CAPI.

SES subgroups also differed by mode of administration. The self-administered web option was completed by 62 percent of highest SES quarter respondents—as opposed to 32 percent for the lowest quarter ($z = 25.68$, $p < .01$).

As expected, more first follow-up nonrespondents were interviewed via CATI/CAPI than self-administered, 70 percent as opposed to 30 percent ($z = 11.34$, $p < .01$). Furthermore, 25 percent of the interviewed cases were conducted via CAPI, which is indicative of the importance of the field option for difficult cases. Similarly, respondents who were offered the "ever dropped out" incentive also were more likely to be interviewed in CATI or CAPI (75 percent, versus 25 percent for web self-administered [$z = 15.64$, $p < .01$]).

4.3.2.2 Incentive Results

As earlier noted, the incentive plan took into account sample member status, timing of interview completion, and degree of case difficulty. Specifically, a higher incentive was offered to sample members who qualified as ever having dropped out, first follow-up nonrespondents, early web respondents, and sample members requiring extra effort to find, reach, or gain cooperation. Sample members were offered incentive amounts ranging from $20 to $60 depending on the above criteria.

Incentive results are presented in two ways. Table 46 provides incentive type by overall interview completions. Table 47 provides incentive/subgroup type by number of cases remaining. Four incentive types are shown: Early, Regular, Difficult, and Final Difficult. Those who completed the survey by web within the first 4 weeks of data collection received the Early incentive, where $10 was added to sample members' base amounts. Once the early completion window had closed, respondents received the Regular incentive which included base incentive amounts only. As data collection efforts continued and case difficulty increased, many sample members became eligible for the Difficult incentive, which once again added $10 to base amounts. By July, all remaining sample members became eligible for a final push, or Final Difficult incentive, which added an additional $10 or $20 depending on respondent type.[59]

Table 46. Interview completions, by incentive type: 2006

Response status and incentive type[1]	Number of cases	Number of completed interviews	Percentage of cases completed
Total	15,900	14,200	89.1
Early	15,900	5,000	31.4
Regular	10,900	5,000	31.4
Difficult	5,900	2,000	12.6
Final Difficult	3,900	2,200	13.6

[1] The Early incentive (base amount plus $10) was offered upon completion by web during the first 4 weeks of data collection. The Regular incentive constituted the base amount. The Difficult incentive added $10 to the base amount. The Final Difficult incentive added an additional $10 or $20 depending on respondent type. See section 4.3.1.2 for actual incentive amounts.

NOTE: Detail may not sum to totals because of rounding. Provided percentages are unweighted. Response rate calculation excludes those cases that are permanently out of scope (deceased) or temporarily out of scope (incapable, unavailable for duration of data collection: e.g., out of the country, incarcerated or institutionalized). Total number of permanently or temporarily out-of-scope second follow-up sample members = 460. Also, unfielded cases are not counted in the second follow-up response rate. In addition to a handful of sample members who asked to be withdrawn from the study, the following in-scope sample members were not fielded in 2006: double (base-year + first follow-up) nonrespondents (n = 330) and first follow-up freshened spring-term senior nonrespondents (n = 40).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

[59] Refer to section 4.3.1.2 for actual incentive amounts.

Table 47. Interview completions, by incentive type and prior response status: 2006

Incentive type and response status[1, 2]	Number of cases	Number of completed interviews	Percentage of cases completed
Total	15,900	14,200	89.1
Early	15,900	5,000	31.4
F1 respondent, ever dropped out	1,000	250	25.8
F1 nonrespondent, dropout	180	10	7.9
F1 respondent, all others	13,700	4,600	33.6
F1 nonrespondent, all others	1,100	140	12.7
Regular	10,900	5,000	45.9
F1 respondent, ever dropped out	720	270	37.0
F1 nonrespondent, dropout	160	40	22.0
F1 respondent, all others	9,100	4,500	49.4
F1 nonrespondent, all others	930	200	21.5
Difficult	5,900	2,000	34.0
F1 respondent, ever dropped out	450	170	37.2
F1 nonrespondent, dropout	130	30	19.5
F1 respondent, all others	4,600	1,700	36.0
F1 nonrespondent, all others	730	160	22.0
Final Difficult	3,900	2,200	55.5
F1 respondent, ever dropped out	280	170	61.3
F1 nonrespondent, dropout	100	40	40.8
F1 respondent, all others	2,900	1,700	58.9
F1 nonrespondent, all others	570	210	37.5

[1] The “early incentive” (base amount plus $10) was offered upon completion by web (or CATI call-in) during the first 4 weeks of data collection. The Regular incentive constituted the base amount. The Difficult incentive added $10 to the base amount. The Final Difficult incentive added an additional $10 or $20 depending on respondent type. See section 4.3.1.2 for actual incentive amounts.

[2] For “ever dropped out”: classified as dropout if at least one of the following conditions was met: school reported that respondent had dropped out of school at any one of the enrollment status updates, respondent was a dropout as of spring term of 2004, or respondent was an alternative completer, that is, earned a GED on or before March 15, 2004.

NOTE: Detail may not sum to totals because of rounding. Provided percentages are unweighted. Response rate calculation excludes those cases that are permanently out of scope (deceased) or temporarily out of scope (incapable, unavailable for duration of second follow-up data collection: out of the country, incarcerated or institutionalized). Total number of permanently or temporarily out of scope second follow-up sample members = 460. Also, unfielded cases are not counted in the second follow-up response rate, which is condition on cases being fielded. In addition to a handful of sample members who asked to be withdrawn from the study, the following in-scope sample members were not fielded in 2006: double (base year + first follow-up) nonrespondents (n=330) and first follow-up freshened senior nonrespondents (n=40). F1 = first follow-up. CATI = computer-assisted telephone interview.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Of the completed interviews, 63 percent of the sample either completed the interview early or during the regular data collection period, while 26 percent completed the interview in the final incentive phases. Interestingly, just as many responded during the initial month of data collection as did those during the regular period from mid-February through June. Almost one third of respondents (31 percent) took advantage of the early web option in the first month of data collection. Respondents, however, were motivated at both ends of the data collection window. When combining the Early and Difficult incentive types, 58 percent of respondents received an incentive that was higher than the base amount offered.

Table 47 provides additional detail across sample member subgroups, including prior response status and “ever dropped out” status. For each subgroup at each incentive level, the number of cases remaining and the number of completed interviews is provided, with a

calculated percent of respondents per row. The incentive strategy was implemented in a series of stages, to balance costs, timing, and methods. Both "nondropouts" (59 percent) and "ever dropped out" qualifiers (61 percent) completed the interview when $50 and $60 were offered, respectively. The $10 sent to sample members in the form of prepayment may conceivably have helped to encourage participation. The smallest gain in cooperation for first follow-up nonrespondents (both nondropouts [13 percent] and ever dropped out qualifiers [8 percent]), came in the Early incentive period. Overall, the Regular and Final Difficult incentive opportunities proved productive, given the total number of cases remaining, 46 and 56 percent, respectively.

4.3.2.3 Process Statistics: Interviewer Effort

Select evaluations of processes related to interviewer effort are provided in this section. In particular, telephone interviewer hours, call counts by response status, and field interviewing results are discussed.

Telephone interviewer hours. The CATI component of data collection required focused effort by telephone interviewers and related staff. The main tasks of contacting and interviewing sample members take many hours, and exclude associated tasks such as training, monitoring, and supervising. Telephone interviewers for the ELS:2002 second follow-up required a total of 20,636 hours, with an average of 3.27 hours spent per completed interview. With an average interview completion time of 27.5 minutes for CATI cases, about 2.8 hours were spent in activities outside the actual interview. The majority of this time was dedicated to locating and contacting efforts. Interviewers were provided multiple contacts per sample member. Interviewers used multiple efforts to locate sample members. Other time was spent on case maintenance, including pulling up a case, reviewing the call history, and closing the case, which may have involved rescheduling an appropriate callback, providing a comment, or updating the case status accordingly.

Number of calls. The majority of interviewer time was dedicated to locating and contacting sample members. This activity requires an extensive outbound calling effort, with some respondent types requiring more calls than others. Table 48 provides call counts by present and prior response status, including counts by mode overall and for second follow-up respondents. About 294,000 calls were made to sample members in the ELS:2002 second follow-up survey. An average of 19 calls were made per case regardless of present or prior response status.

Table 48. Call counts, by present and prior response status: 2006

Response status	Overall		
	Number of cases	Total number of calls	Average calls per case
Total	15,900	293,900	18.5
F2 respondent			
F1 respondent	13,300	195,400	14.7
F1 nonrespondent	800	18,100	21.9
F2 nonrespondent			
F1 respondent	1,300	66,900	50.8
F1 nonrespondent	400	13,600	32.6

NOTE: Detail may not sum to totals because of rounding. Response rate calculation excludes those cases that are permanently out of scope (deceased) or temporarily out of scope (incapable, unavailable for duration of second follow-up data collection: e.g., out of the country, incarcerated or institutionalized). Total number of permanently or temporarily out-of-scope second follow-up sample members = 460. Also, unfielded cases are not counted in the second follow-up response rate, which is conditional on a case being fielded. In addition to a handful of sample members who asked to be withdrawn from the study, the following in-scope sample members were not fielded in 2006: double (base year + first follow-up) nonrespondents (n = 330) and first follow-up freshened senior nonrespondents (n = 40). F1 = first follow-up; F2 = second follow-up.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Of those who completed the 2006 interview, first follow-up respondents received an average of 15 calls and first follow-up nonrespondents received an average of 22 calls. A concentrated effort was made to survey second follow-up nonrespondents. In particular, those who had responded in the first follow-up but declined to participate in the second follow-up were called an average of 51 times. Conversely, those who did not respond in the first follow-up were called an average of 33 times per case.

Field interviewing. Field interviewers were able to pursue contacting efforts both by telephone and in person, and had advantages including local area calling and face-to-face interaction with sample members. With this enhanced accessibility, interviewers employing CAPI efforts are often able to secure participation when other data collection efforts are not successful.

Table 49 presents response rates by field interviewing status. The majority of the sample (81 percent) was successfully interviewed without field follow-up. However, more than 1,330 cases (8 percent of the sample) were successfully interviewed only with the help of field follow-up. Of all cases sent to the field, 57 percent were successfully interviewed and 43 percent were not. Some outstanding cases (730) were not sent to the field owing to such factors as firm refusal or unusable contact information.

Table 49. Interview completion and noncompletion rates, by field status: 2006

Case type	Number of cases	Percent
Total	15,900	100.0
Interviewed without field follow-up needed	12,800	80.7
Interviewed, field follow-up required	1,300	8.4
Not interviewed, field follow-up attempted	1,000	6.3
Not interviewed, no field follow-up attempted	730	4.6

NOTE: Provided percentages are unweighted. Detail may not sum to totals because of rounding. Response rate calculation excludes those cases that are permanently out of scope (deceased, sampling error) or temporarily out of scope (incapable, unavailable for duration of data collection: e.g., out of the country, incarcerated, institutionalized). Total number out of scope = 460. Also, unfielded cases are not counted in the second follow-up response rate. In addition to a handful of sample members who asked to be withdrawn from the study, the following in-scope sample members were not fielded in 2006: double (base-year + first follow-up) nonrespondents (n = 330) and first follow-up freshened spring-term senior nonrespondents (n = 40).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

4.3.2.4 Evaluation of Data Quality

This section includes select evaluations of the quality of data collected in the 2006 interview. Interview completion time is discussed overall, by questionnaire section, and by mode (web, CATI, or CAPI). Coding systems within the instrument for field of study and occupation are also discussed, including an analysis of coding accuracy. Telephone interviewer performance in question delivery and data entry is also assessed.

Interview completion time. To calculate the time required to complete the survey, start and end time stamp variables were associated with each question. Time stamps were recorded using the respondent's or interviewer's computer clock time. As respondents or interviewers moved from screen to screen, actual on-screen times and transit times between screens were recorded and summed. Section times and total instrument times were then calculated accordingly.

On average, respondents took about 27 minutes to complete the ELS:2002 second follow-up survey. Table 50 shows average completion time overall and by section, and mode of administration.

Table 50. Average minutes to complete interview, by interview section and mode: 2006

Instrument section	Respondents			
	All	Web	CATI	CAPI
Total interview	27.2	26.5	27.5	28.8
Section A—High school	1.5	1.2	1.5	2.3
Section B—Postsecondary	11.0	11.7	10.6	8.9
Section C—Employment	7.1	6.5	7.3	8.8
Section D—Community	3.3	3.1	3.4	3.3
Section E—Locating	5.5	4.9	5.8	6.8

NOTE: Outliers were excluded from analysis. An outlier was defined as any question requiring more than 5 minutes' response time. Interview times are based on completed interviews only. Abbreviated English, Spanish, and partial cases were excluded from analysis. CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Section B (postsecondary education) had the longest average completion time (11 minutes). Section A (high school) was inapplicable for most respondents, lowering the average time to less than 2 minutes.

Section C (employment) showed the second longest time, 7 minutes. Most respondents, in particular standard enrollees, were asked limited questions about employment due to being currently enrolled. Section D (community) was a relatively quicker section, taking about 3 minutes to complete. The last section, Section E (locating), took over 5 minutes to collect contact information for future follow-up.

Analysis of field of study and occupation recoding. The ELS:2002 second follow-up instrument enabled sample members and interviewers to code verbatim responses given for field of study and occupation. Currently enrolled respondents were asked for their field of study, while all respondents were asked about the job they expected to have at age 30. A subset of respondents, depending on their enrollment status and history, was asked about first job after high school and/or current job.

Both coding systems used an assisted-coding approach. Entered text describing the field of study or occupation interfaced with a database to provide a candidate for the best category match or provide a set of comparable matches. The assisted-coding approach—as opposed to search-and-select or manual coding—presents less burden on the respondent and interviewer, requiring less time to code, and streamlining the selection process. If the system could not make a match (e.g., owing to misspelling) or if a selection could not be made from the list displayed, respondents and interviewers were routed to a double or triple dropdown screen to make a selection manually.

For field of study, the category names provided by assisted coding were synonymous with the general and specific categories provided by the manual dropdowns. The field of study coder provided 33 general categories and 192 specific categories. The categorical framework was largely based on the most recent version of the Classification of Instructional Programs (CIP-2000), which provides a taxonomy of instructional program classifications and descriptions.

The occupation coding system used O*NET (Occupational Information Network; http://online.onetcenter.org/). The O*NET database was developed for the U.S. Department of Labor and represents an extensive set of worker attributes and job characteristics. O*NET provides a nested coding scheme; 23 general-level categories expand to 96 midlevel categories, which expand to 821 specific-level categories.[60] Specific level occupations can therefore roll up to broader categorizations. If an occupational match could not be found using assisted coding, a triple dropdown menu enabled manual selection. For job expected at age 30, however, the manual coder was bypassed altogether. This was done due to the hypothetical nature of asking about a future job that may be more difficult to assign a specific code. Any verbatim responses that were not coded during the interview were coded by expert staff after data collection.

To assess the reliability of coding procedures, two occupational coding specialists evaluated random samples of coded responses. Ten percent of field of study responses and 10 percent from each occupation variable were assessed for coding accuracy. Table 51 shows the

[60] Refer to appendix F for a detailed occupational crosswalk providing all O*NET classifications (general, midlevel, specific), in addition to their paired mappings to the original 16 occupational categories used in the base year and first follow-up of ELS:2002.

results of the recode analysis, including the number of responses sampled by mode, accuracy of the original code, and the percentage of strings too vague for recoding. Assessment of coding accuracy is based on the specific level of coding, meaning responses were deemed correct or incorrect at the most specific category level for both field of study and occupation.

Table 51. Summary of recode results, by mode: 2006

Type of coding	Web respondents			CATI/CAPI respondents		
	Coding attempts sampled	Percent original code correct	Percent text string too vague to code	Coding attempts sampled	Percent original code correct	Percent text string too vague to code
Total	940	73.5	1.3	1,400	82.0	1.9
Major/field of study	390	74.5	0.3	240	86.4	0.0
Occupation	550	72.8	2.0	1,100	81.0	2.3
First job after high school	130	65.6	1.6	290	77.2	3.5
Current job	90	75.0	2.3	210	80.8	1.4
Job expected at age 30	330	74.9	2.1	620	82.9	2.1

NOTE: Detail may not sum to totals because of rounding. Provided percentages are unweighted. CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Overall, both coding systems fared well in usability and accuracy, lending support for the use of an assisted-coding approach. Coding accuracy generally ranged from 77 to 86 percent for CATI/CAPI interviews and 66 to 75 percent for web respondents. As expected, interviewer-administered cases showed higher accuracy overall compared to self-administered cases—82 percent as opposed to 74 percent ($z = 4.87$, $p < .01$). Specifically, interviewers correctly coded field of study at a higher rate than web respondents—86 percent as opposed to 75 percent ($z = 3.58$, $p < .01$). Interviewers also coded occupation more accurately—81 percent as opposed to 73 percent ($z = 3.81$, $p < .01$). All interviewers were trained in using the coding systems, and became familiar with the mechanics of coding to minimize coding time during the interview. Web respondents, however, were provided with on-screen brief instructions to assist with coding. Any originally incorrect responses were recoded accordingly (18 percent of CATI/CAPI responses and 26 percent of web responses), in order to reflect accurate field of study or occupation categorizations and improve data quality.

Given the structure of O*NET, coding accuracy can be assessed at three levels of detail (general, midlevel, specific). Specific-level codes can roll up to midlevel and general-level categories. The 23 general-level categories within O*NET represent a comprehensive and manageable set of contemporary job categories. When assessing coding reliability at the general O*NET level, as one would expect, matches were greater at more general levels. Excluding the small number of cases deemed too vague for recoding purposes, occupation at the general level was coded correctly 88 percent of the time overall. Interviewer-administered cases showed higher accuracy compared to self-administered—90 percent as opposed to 84 percent ($z = 3.38$, $p < .01$). For both modes, this is a large gain in accuracy at the level containing 23 categories; an 8 percent increase for CATI/CAPI and a 10 percent increase for web respondents.

Further assessment of coding accuracy can be seen in the mapping of O*NET codes to ELS:2002 occupation codes. The base-year and first follow-up rounds of ELS:2002 used an occupation coding scheme consistent with predecessor studies, including NELS:88. To use the efficiency of O*NET, yet also provide consistency with previous rounds, all 821 specific-level

O*NET codes were mapped accordingly to the ELS:2002 occupation coding schema of 16 categories.[61] Given this crosswalk, coding accuracy to the ELS:2002 schematic can also be assessed. With the exclusion of vague responses, associated ELS:2002 categories were coded correctly 85 percent of the time overall. CATI/CAPI also had the advantage compared to web, with 88 percent correct as opposed to 80 percent ($z = 4.27$, $p < .01$).

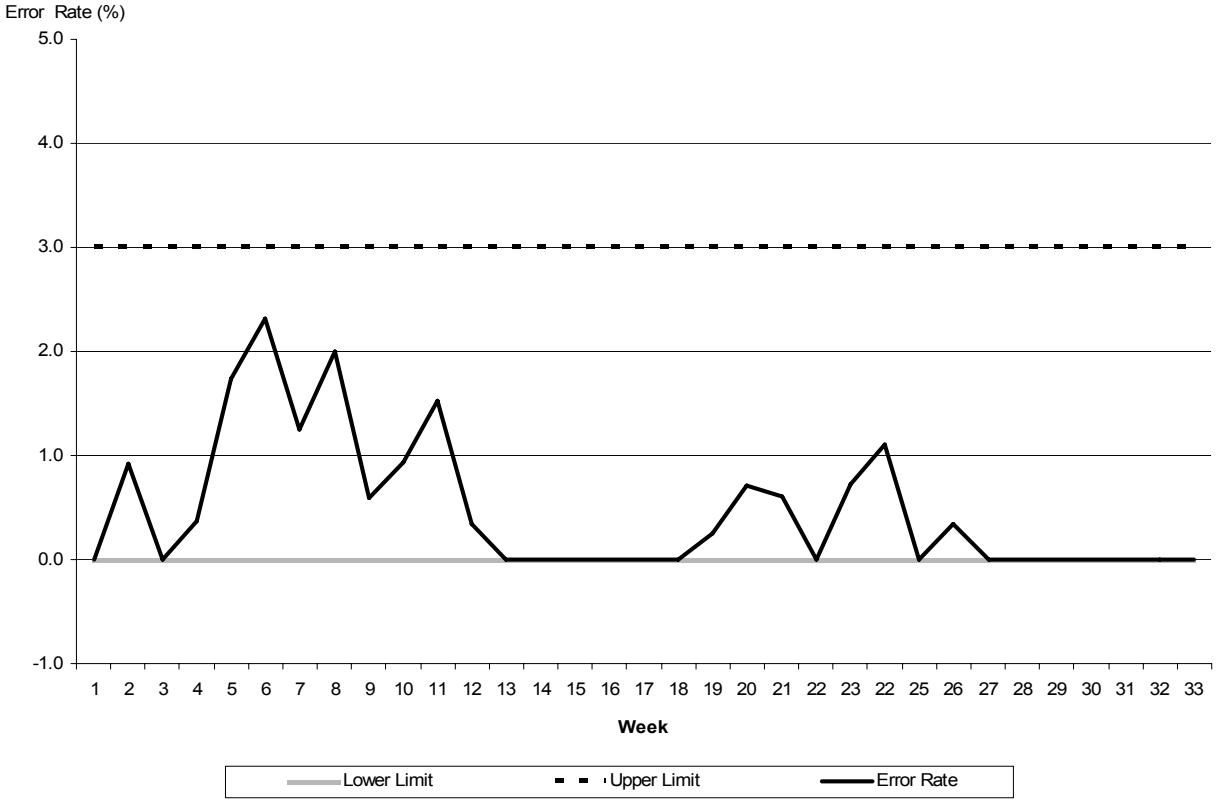
Question delivery and data entry error rates. CATI interviews were regularly monitored throughout data collection, from late January through early September. Monitoring helps improve interviewing and enhances data quality. For studies with an interviewer component, ensuring both standardized interview delivery and appropriate data capture is important. Monitoring helps to meet the following objectives: identify problematic items, reduce interviewer error, improve interviewer performance by reinforcing procedures and strategies, and assess the quality of data collected.

Interviewer performance was evaluated in two ways: (1) how interviewers administered items to the respondent and (2) how interviewers recorded responses. Specially trained monitors were able to concurrently view and listen to live CATI interviews without disturbing the interviewer or respondent. Monitoring equipment facilitated remote observation with the flexibility to tune into any interview. Monitors observed blocks of up to 20 questions per interview, and evaluations were conducted during all shifts, including day, evening, and weekend.

During CATI data collection, 9,885 items were monitored. Of these, monitoring staff observed 89 total errors, yielding an overall error rates of just 0.9 percent. Three percent was defined as the boundary for the weekly error rate, above which direct intervention would be required. Question delivery incurred 71 errors (a 0.7 percent error rate; data entry incurred 18 errors [an error rate of 0.2 percent]). Question delivery errors and data entry errors are illustrated in figures 5 and 6, respectively. Typically, weekly error rates fell below 2.3 percent. Most fell below 1 percent; many weeks showed no errors at all. The peaks in error rate are attributable to the addition of new interviewer staff, who are more prone to errors due to inexperience. Monitoring efforts were the most intensive early on in data collection, particularly after the first major CATI interviewer training, from late February through late March. By early August, monitoring efforts were reduced given the lighter caseload and consistently low error rates.

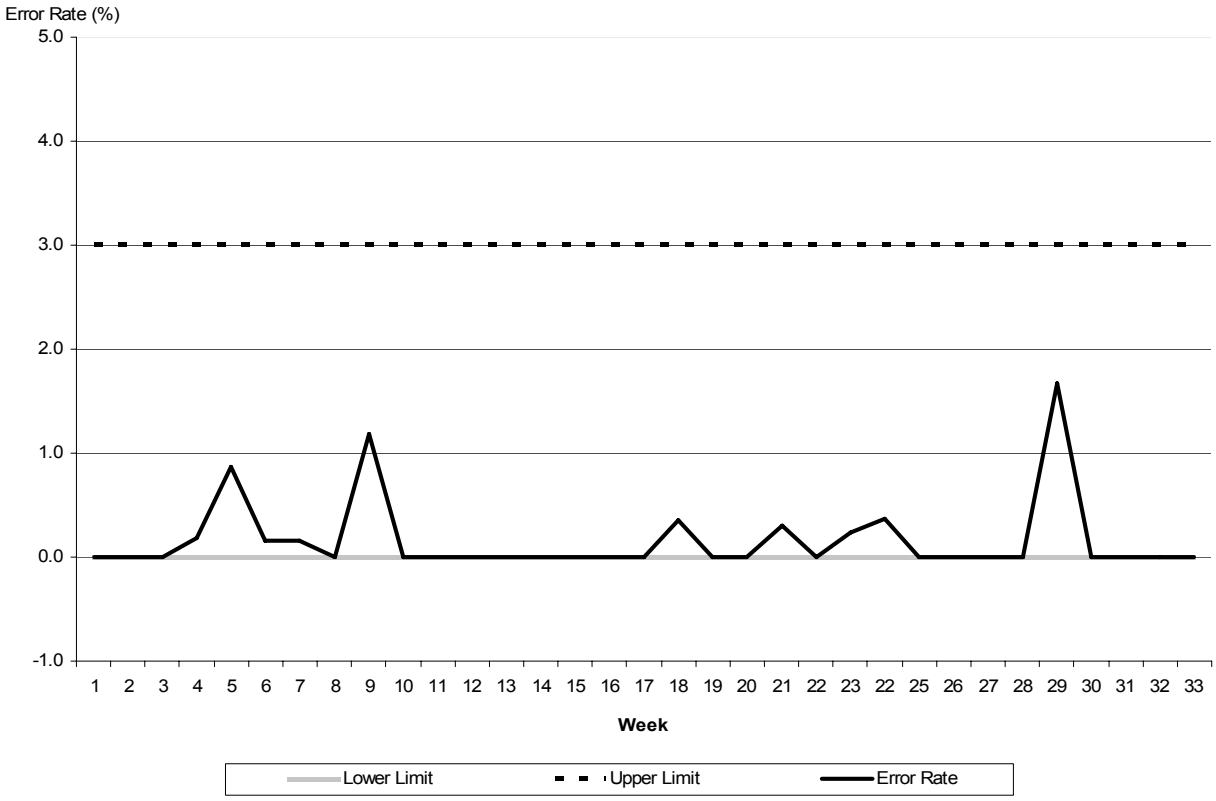
[61] Refer to appendix F for a detailed occupational crosswalk providing all O*NET classifications (general, midlevel, specific), in addition to their paired mappings to the original 16 occupational categories used in the base year and first follow-up of ELS:2002.

Figure 5. ELS:2002 second follow-up quality assurance monitoring results by week for question delivery error rates: 2006



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Figure 6. ELS:2002 second follow-up quality assurance monitoring results by week for data entry error rates: 2006



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Chapter 5
Data Preparation and Processing

5.1 Base-Year and First Follow-up Data Preparation and Processing

This chapter describes the automated systems used to control survey processes for the Education Longitudinal Study of 2002 (ELS:2002), including procedures used to maintain receipt control; aspects of data preparation (such as coding); and the various procedures for data capture, cleaning, and editing. The discussion in this chapter includes data obtained from questionnaires, academic records (transcripts and course catalogues), and all other sources.

5.1.1 Overview of Systems Design, Development, and Testing

Most systems were developed in the base year, then redesigned if necessary during the first follow-up field test with concern for the processes needed for the first follow-up main study. The effort was to test systems in a smaller environment to reveal points in which improvements could be implemented on a larger scale. After the field test, improvements were implemented and checked in a test environment.

The following systems were developed in the base year and refined and tested in the first follow-up field test:

- a recruiting system;
- a Survey Control System (SCS);
- a Survey Day materials generation program;
- a questionnaire receipt application;
- a web-based Integrated Management System;
- production reports;
- Teleform (application used for scanning questionnaires);
- a mail return application;
- an incentive tracking application;
- a field reporting system to help field supervisors track the status of in-school data collection and field interviewing;
- a Structured Query Language (SQL) server database to store scanned data responses;
- a scanned image database; and
- a student computer-assisted telephone interview (CATI) instrument.

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

and were updated to reflect what changed between the field test questionnaires and the full-scale questionnaires.

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

- a field assignment system;
- a field materials generation system;
- mail generation invoked by requests in CATI;
- a computer-assisted data entry program for the field screener;
- Teleform versions of out-of-school hardcopy questionnaires (i.e., transfer, dropout, early graduate);
- quality control steps implemented during scanning, rather than later during data delivery processes;
- data cleaning and editing programs;
- a scanned image archive server that allowed instant access to scanned questionnaires during the data cleaning and review process;
- a cleaning and editing application that allowed editors to review and correct questionnaire data as appropriate, working in conjunction with actual scanned images in cases in which inconsistent data occurred;
- a data review system that allowed reviewers to randomly review questionnaires with data to detect data deficiencies (e.g., scanning problems); and
- an occupation coding application.

5.1.2 Base-Year and First Follow-up Data Receipt

The data preparation facility received all materials returned to RTI after a school's survey was complete or school officials sent in completed questionnaires. Procedures were established to systematically receive and record all required forms; this process included the scanning of bar-coded labels. Receipt events were available for the full-scale study to identify questionnaires that were not completed fully or accurately and to allow project staff to follow up promptly. Different versions of questionnaires (e.g., student, transfer, early graduate, etc.) were easily distinguishable within the receipt process and were automatically batched separately based on the questionnaire type.

After questionnaires were received and added to the receipt system, a batch number was assigned to the questionnaire. To assist the project team in cases that required referring to a questionnaire, the system was able to access dynamically the status of an individual questionnaire and provide its batch number. If the questionnaire had moved beyond the scanning stage, the scanned image could be accessed as well. Questionnaires were occasionally identified for data removal (e.g., when parental consent was lacking). Rather than deal with the removal process manually, a spreadsheet was developed to document these cases, and case removal was integrated into the data delivery process. This approach was useful because it did not disrupt the

questionnaire processes and provided the ability to add cases back to final data files when appropriate (e.g., when parental permission was obtained).

5.1.3 Occupation Coding for Hardcopy Instruments

In the base year, occupation was coded from text in the parent and student questionnaires. In the first follow-up, occupation was coded from the student questionnaire and new participant supplement. Occupation text was loaded into a coding application in which a coding specialist could select the correct code from the 16 occupation categories. The resulting codes were merged back into the data files.

5.1.4 Base-Year and First Follow-up Data Capture for Scanned Instruments

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

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

5.1.5 Base-Year and First Follow-up Cleaning and Editing for Hardcopy Questionnaire Data

An application was developed in which case/item-specific issues were reviewed and new values were recorded for subsequent data cleaning and editing. Records were selected for review based on one of the following criteria: random selection, suspicious values during frequency reviews, values out of expected ranges, and values not adhering to a particular skip pattern. The review application provided the case/item-level information, reasons for review, and a link to the scanned image of the questionnaire. Reviewers determined scanning corrections, recommended changes (if respondents had misinterpreted the question), and reviewed items randomly to spot potential problems that would require more widespread review.

The application was built on an SQL server database that contained all records for review and stored the recommended data changes. Editing programs built in SAS read the SQL server database to obtain the edits and applied the edits to the questionnaire data. Questionnaire data

were stored at multiple stages across cleaning and editing programs, so comparison across each stage of data cleaning could be easily confirmed with the documentation on recommended edits. Raw data were never directly updated, so changes were always stored cumulatively and applied each time a cleaned dataset was produced. This process provided the ability to document all changes and easily fix errors or reverse decisions upon further review.

Editing programs also contained procedures that output inconsistent items across logical patterns within the questionnaire. For example, instructions to skip items could be based on previously answered questions; however, the respondent may not have followed the proper pattern based on the previous answers. These items were reviewed, and rules were written to either correct previously answered (or unanswered) questions to match the dependent items or blank out subsequent items to stay consistent with previously answered items.

5.1.6 Base-Year and First Follow-up Data Capture and Editing for CATI

In the base year, a CATI version of the parent questionnaire was employed. In the first follow-up, for the out-of-school data collection effort, the following CATI instruments were developed to administer to sample members: student (developed from the Teleform abbreviated version), transfer, not currently in school (dropout), early graduate, and homeschool. A screener at the beginning of the CATI survey was responsible for determining which questionnaire module a respondent was to be administered.

CATI logic was designed such that the Teleform and CATI records could be concatenated into one data file. CATI instruments were developed with logic based on the skip patterns in the questionnaires. Questions were automatically skipped during administration. The questionnaire development program (Blaise) stored data for each item answered, but respondents were allowed to go back to previously answered items. In rare cases, a previously answered item could be changed in such a way that the questionnaire logic was inconsistent with data already answered from a different logical path. Blaise automatically corrected the previously administered responses so that the skip logic was consistent.

5.1.7 Base-Year and First Follow-up Data Processing and File Preparation

All Teleform questionnaire scans were stored in an SQL server database. CATI data were exported nightly to ASCII files. Cleaning programs were designed to concatenate CATI and Teleform SQL server data into SAS datasets, adjusting and cleaning variables when formats were not consistent. Special attention was focused on this concatenation to verify that results stayed consistent and to rule out possible format problems.

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

- anomalous data cleaning based on review of data with original questionnaire image (e.g., scanning errors);
- rule-based cleaning (changes that were made based on patterns in data, rather than review of images);
- hard-coded edits based on changes recommended by a reviewer if respondents misunderstood the questionnaire (e.g., respondent was instructed to enter a

percentage, but there was strong evidence that the respondent entered a count instead); and

- edits based on logical patterns in questionnaire (e.g., skip pattern relationships between gate and dependent questions).

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

Item documentation procedures were developed to capture variable and value labels for each item. Item wording for each question was also provided as part of the documentation. This information was loaded into a documentation database that could export final data file layouts and format statements used to produce formatted frequencies for review. The documentation database also had tools to produce final electronic codebook input files.

5.2 First Follow-up Transcript and Course Offerings Procedures

This section summarizes procedures associated with the processing of high school academic transcripts and course catalogs. For detailed information on archival records collection and processing based on student transcripts and high school course catalogs, see Bozick et al. (2006), which is available only with the restricted-use transcript files. The Institute of Education Sciences/National Center for Education Statistics will only accept restricted-use data license applications through its electronic application system (see http://nces.ed.gov/statprog/instruct.asp). More information about applying for restricted-use data licenses is available at http://nces.ed.gov/statprog/instruct.asp and in the "Restricted-Use Data Procedures Manual" at http://nces.ed.gov/statprog/rudman/toc.asp.

A concise introduction to the transcript data is provided by Planty, Bozick, and Ingels (2006).

5.2.1 First Follow-up Transcript Procedures

5.2.1.1 Receipt Control

Incoming data collection forms, transcripts, and course catalogs were logged into the survey control system by staff in RTI's data preparation unit. Data editors reviewed each school's packet of materials for completeness and legibility. Data editors first recorded whether the Transcript Cover Sheet and Student Transcript Checklist were completed and returned by the school. The Transcript Cover Sheet was examined to determine if any of the requested items were unavailable, and this information was recorded in the survey control system. Data entry clerks keyed the data from Transcript Cover Sheet and Student Transcript Checklist forms in the survey control system. Assigned institutional contactors (ICs) called schools to follow up regarding any missing materials. Missing materials were retrieved by telephone or mail. The results of each school contact were recorded in the survey control system.

At the student level, individual transcript receipts were recorded in the survey control system by data preparation staff. Once the items were recorded, data editors reviewed them for legibility and completeness. Packets with edit problems were routed to a supervisor for

resolution. Electronic reports were produced and monitored to identify missing or unclear information at the school and student levels. ICs followed up to obtain missing documents and to clarify information on the student transcripts. Items that were both legible and complete were routed to keying and coding.

5.2.1.2 Course Catalog and Transcript Entry

Course catalog and transcript data were entered using a web-based, computer-assisted data entry system. This system consisted of sequential data entry screens grouped by type of information requested (school-, student-, or course-level data). Identifying information such as identification number, school name, and student names were preloaded into the data entry system. Quality checks such as valid ranges, data types (e.g., numeric or character), and field sizes were specified for each data element; keyer-coders were required to reenter data failing these checks. Keyer-coders were responsible for keying school-, student-, and course-level data and for coding course data. A quality control team verified all keyed data. A supervisor and a team of experienced keyer-coders were on site at all times to manage the effort and provide guidance when needed.

5.2.1.3 Course Catalog Entry

Course catalogs from ELS:2002 base-year schools were keyed and coded for the preparation of course offerings data. Only course offering information for base-year schools appears on the course-level file. While catalogues were collected for up to four academic years, whenever possible a school's 2003–04 course catalog was used. Each school was assigned to a single keyer-coder for course catalog entry. Information entered included the following:

- School-level information:
  - catalog type and year;
  - term system;
  - grading system;
  - credits equal to one Carnegie unit (schools were asked how many credits a student would earn for taking a course that meets every day, one period a day, all school-year long); and
  - credits required for each type of diploma.
- Course-level information:
  - course name, school-assigned course number, course department name;
  - state/district-assigned course number;
  - credits offered;
  - program type;
  - term(s) course offered;
  - restricted enrollment, if applicable;

  - grade level(s) to which course is offered; and
  - Classification of Secondary School Courses (CSSC) code (see section 5.2.1.4).

The data entry system included a mechanism for setting the status of each school catalog, such as "assigned for keying/coding" and "quality control needed." System-generated reports based on these statuses were used by project staff to monitor progress and to review/edit when necessary.

All transcripts received from any one school were assigned to a single keyer-coder for both student- and course-level data entry. Keyer-coders thoroughly reviewed transcripts and all related materials (e.g., Student Transcript Checklist, Transcript Cover Sheet, and course catalogs) before abstracting data. The Student Transcript Checklist was helpful in providing school-reported student-level data, such as participation in special programs. The following information was entered:

- Student-level information:
  - Participation in specialized programs.
  - Date sample member left school—the graduation or final withdrawal date was entered. Keyer-coders also entered the date the student rejoined the school, if applicable.
  - Reason sample member left school (e.g., graduated or transferred).
  - Type of diploma or equivalency certification received (e.g., standard, honors, or General Educational Development).
  - Cumulative grade point average (GPA), weighted and unweighted—the GPA was entered as reported by the school. When a transcript provided a GPA but did not specify whether it was weighted or unweighted, it was entered as unweighted.
  - Preliminary Scholastic Aptitude Test, Scholastic Aptitude Test (SAT), ACT, Advanced Placement (AP), and/or SAT subject test scores and date taken—the data entry system allowed for multiple test score entries per test type.
- Coursetaking histories:
  - Course name and school-assigned course number—course titles were keyed verbatim, except for the use of approved abbreviations and the conversion of Roman numerals to Arabic. When available, school-assigned course numbers were entered as separate data elements.
  - School year in which the course was taken.
  - Grade level (grade in which the sample member was enrolled at the time the course was taken).
  - School where the course was taken.
  - Term when the course was taken.
  - Credits received (number of credits awarded for the course as reported on the transcript).

- Raw grade (grade received for the course as reported on the transcript).
- Grade received—a standardized letter grade was entered, converted from the raw grade based on the school's grading scale.
- CSSC code (see section 5.2.1.4).

School transcripts provided coursetaking histories at the year or term level. Year-long courses might be reported with a distinct listing (and separate grade) for each term in that school year. For example, a year-long algebra course might appear on a transcript twice, once for fall semester and once for spring semester. When the transcript reported a final (year-end) grade, the course was entered as a year-long course, along with the grade received. When no final (year-end) grade was reported, the course was entered as two semester-long courses, each with the corresponding grade received.

5.2.1.4 Course Catalog and Transcript Course Coding

The CSSC, updated from the 2000 National Assessment of Education Progress high school transcript study, was used for coding all ELS:2002 catalog and transcript courses. The CSSC is designed to describe course offerings in secondary education and to provide a coherent means for classifying these courses. Each CSSC code comprises six digits, with an associated course title, alternate titles, and a course description. The first two digits identify the main program area (e.g., mathematics), the second set of two digits represents a subcategory of courses within the main program area (e.g., pure mathematics), and the last two digits are associated with the specific courses in each of the main and subcategories (e.g., trigonometry).

For ELS:2002 base-year schools that provided them, courses listed in course catalogs were keyed and assigned the appropriate CSSC code before transcript keying and coding. This order of procedures enhanced the quality and consistency of the coding process. Then, transcript courses could be accurately coded by simply matching their titles with the titles of courses in the course catalogs. Otherwise, each course on the transcripts would have to be matched one by one to a CSSC code based only on the course title and the CSSC course title, with none of the information describing the course content usually included in the school's course catalog.

For each catalog course entered, keyer-coders selected an appropriate course code from the CSSC look-up table in the data entry system. The look-up table included CSSC course codes, titles, and descriptions. Keyer-coders could search course codes by course title, description, keywords, or a combination of these. Using the look-up table in the system reduced hardcopy look-up time. The CSSC code was selected after reviewing the course description and any relevant school-level information from the course catalog. The data entry system checked the validity of each selected CSSC code before accepting it. To further increase coding efficiency, RTI developed a subset of frequently used CSSC codes. This list was also available as a look-up table in the data entry system and was expanded and maintained throughout the coding process. Because of changes in the curriculum, a handful of "new" courses were identified and assigned new CSSC codes.

All transcripts received from a school were assigned to a single person for keying and coding. Each sample member's courses were coded individually. For ELS:2002 base-year schools that provided transcripts and a course catalog, transcript coding took place after that school's catalog had been coded and keyed. Coding consistency and speed were increased

because the data entry system allowed keyer-coders to select CSSC codes for transcript courses by matching them with corresponding catalog courses. When prompted for a transcript course code, keyer-coders were supplied with a list of all courses keyed from the school's catalog. Keyer-coders could browse the entire list, or search by course name or course number. Upon selecting a matching catalog course, the keyer-coders could assign the catalog course's code to the transcript course. If the keyer-coders could not find an acceptable match, a CSSC code was selected from the master CSSC list. If no CSSC code was deemed appropriate, the keyer-coders marked the course as uncodeable (600000).

Course catalogs from non-base-year schools were not keyed. These schools' transcript courses were coded using the school-provided course catalog as a resource to provide a course description, an overview of the school curriculum, and other valuable information. The keyer-coder used the look-up table to select the appropriate CSSC code, and the data entry system checked the validity of each CSSC code before accepting it.

Of the 1,557 schools that provided transcripts, only 24 (2 percent) did not provide a catalog. When possible, a substitute catalog was identified from the pool of sampled schools that provided one to use as a resource for coding. Substitute catalogs were selected from schools in the same district (or state, if necessary) and on the basis of size and type (public or private; and school affiliation, where applicable). Keyer-coders then used the substitute catalog as a resource for coding transcript courses. In rare cases where no suitable substitute catalog was available, transcript courses were coded according to course title, grade level, course level, and track indicators.

Data entry of each catalog and transcript was reviewed for accuracy by a supervisor or by a group of keyer-coders trained to perform these reviews. Any inconsistencies between the source document and corresponding data entered were corrected. The data entry system recorded the corrected errors and calculated error rates for each keyer-coder. Those with high error rates were identified and retrained as necessary.

Quality control of course entry and coding involved several components. First, preliminary work performed by each newly trained keyer-coder was reviewed. After a hands-on examination of source documents and selected codes, a coding supervisor met with each keyer-coder individually to provide feedback and to make corrections. Individual guidance continued, if necessary, until the keyer-coder reached an acceptable level of independence and coding mastery.

Course coding was reviewed by expert coders in several key areas: coding of AP courses, coding of special education courses, coding consistency within schools, and accurate coding based on track and sequence indicators. When the expert coder disagreed with a code assigned by a keyer-coder, the code was changed in the data entry system. In addition, all catalog and transcript courses marked as uncodeable were reviewed. CSSC codes were applied where possible, including the use of recommended new codes. Unusual course abbreviations (a more common problem with transcripts than catalogs) were investigated, deciphered, and coded wherever possible. A small percentage of nondescript courses such as "Mini-course" or "Transfer Elective" were left as uncodeable (600000), despite all efforts to determine an appropriate code. Of the total transcript courses, 1 percent were uncodeable. Lastly, keyers and coders inspected all student-level records to ensure that there was no duplicated information in the data file resulting from multiple transcripts.

5.2.1.5 Machine Edit

Procedures for editing, coding, error resolution, and documentation were modeled after the National Education Longitudinal Study of 1988 second follow-up transcript component (Ingels et al. 1995). Data entry systems included valid ranges and codes, including legitimate missing codes, and CSSC code checks. Sequences of machine edits and visual data inspections were performed. Tasks included supplying missing data, detecting and correcting illegal codes, and investigating and resolving inconsistencies or anomalies in the data. Variable frequencies and cross-tabulations were reviewed to verify the correctness of machine editing.

After all improperly entered data were corrected, the transcript data passed through a second step in the editing program that supplied the appropriate reserve codes to fill blank fields. The reserve codes are as follows: -4: Nonrespondent, and -9: Missing.

Transcripts were received and systematically entered in the survey control system. They were then tracked as they continued through coding procedures. Once all transcript keying and coding was completed, the following cleaning and editing steps were implemented:

- cleaning anomalous data based on review of data with original transcripts (e.g., keying errors);
- removing duplicate course data erroneously provided by schools on the transcripts or duplicated across school transcripts;
- converting course credits to Carnegie units based on a school conversion factor;
- supplementing transcript information captured in the survey control system when information was missing on transcripts; and
- applying appropriate reserve codes where information was not available.

Next, the following records were examined individually because they indicate potentially anomalous and/or unlikely academic situations:

- all courses in schools where at least one student earned more than 35 Carnegie units;
- all courses in schools where at least one student earned less than 20 Carnegie units;
- courses associated with students who earned more than two Carnegie units for a single course;
- courses associated with students who earned more than typical Carnegie units in a course and/or subject area (e.g., more than four Carnegie units in math);
- courses associated with students where patterns of grade and academic year were inconsistent (e.g., grade-level changes within year or year changes within grade level; grade levels spanning 2 academic years);
- courses associated with students who had completed high school on time, had complete transcript information, and yet had a GPA of 0.00;
- courses that have passing grades (greater than F) and yet have zero credit; and
- courses taken during terms after the transcript indicates that the student had left high school.

All of these records were examined and corrected when errors were detected. Additionally, all course records with "Advanced Placement," "AP," "International Baccalaureate," or "IB" in the title or courses with AP/IB CSSC codes were examined to ensure that they were adequately identified and coded.

Once these quality control measures were implemented, student-level variables (e.g., graduation status, credits earned in a subject area) were merged onto the existing student file. The student's course information was used to create a new student course file. This file contains multiple records for each student and can be linked back to the student file.

The same cleaning and editing procedure applied to the course catalogs. A school course offerings file was produced for the base-year schools only and provides course information that can be linked to the student course file.

Transcript information was added to the first follow-up restricted-use electronic codebook (ECB) by

- merging student-level transcript information to the student file in a transcript composite section;
- appending new transfer schools to the school file to be linked with student-level and student-course-level transcript information; and
- creating new files for student-course-level data and course offerings data.

Item documentation was created for the transcript variables and files. The first follow-up ECB was extended and includes the following files:

- HSTRNSTU.PRI: Course-level file;
- BYF1TSTU.PRI: Student-level file;
- BYF1TSCH.PRI: School-level file; and
- HSTRNSCH.PRI: Course-offering file.

5.3 Second Follow-up Data Cleaning, Coding, and Editing

A database was developed in which case/item-specific issues were reviewed and new values were recorded for subsequent data cleaning and editing. Records were selected for review based on one of the following criteria: suspicious values during frequency reviews, values out of expected ranges, interviewer remarks, and values not adhering to a particular skip pattern.

In the second follow-up, coding was not part of the post-data collection activities, but took place in the interview itself through self- or interviewer coding (from verbatims) of field of study and occupation, using an automated assisted coding approach. An assessment of coding quality has been provided in chapter 4. The coding scheme used in the second follow-up was taken from O*NET. This scheme is documented in appendix F of this document, and includes a crosswalk to other occupational classification schemes used in ELS:2002.

Editing programs contained procedures that output inconsistent items across logical patterns within the interview. The interview was developed as a web-based instrument available to field interviewers, telephone interviewers, and as a web-based self administered questionnaire (SAQ) for the respondent. The instrument administers a questionnaire based on skip logic. Items

that are dependent on other items are only administered when the skip logic so indicates. The instrument allows the interviewer or SAQ respondent to back up in order to correct responses; however, the instrument leaves data for items that no longer apply. A SAS programmer uses the instrument specifications and programming code to step through the programs and determine where logical patterns and consistencies should be edited, and enters edit statements into a SAS program. Final edited data were passed through the original program to confirm that no item inconsistencies exist.

Items that were related based on data consistencies, but for which a consistency check was not built into the instrument programming, were checked with SAS programs in a post-processing step. Crosstabulations were developed to review logical consistencies across items. Values for items that are input to the crosstabulation were collapsed into similar values to make the crosstabulation more "readable." The crosstabulations were reviewed as a quality control check to determine if there were programmatic errors in cleaning or editing steps. To confirm that data editing and cleaning programs were applying changes appropriately, the following steps were implemented:

- Ran comparisons between raw data and cleaned data and reviewed the results to ensure that they were as expected.
- Reviewed crosstabulations while following the instrument and source code from the final instruments.
- Reviewed frequencies to confirm that values followed an expected pattern.
- Ran frequencies by respondent type.
- Ran SAS editing programs with a temporary step that flags values to be blanked out and allows for review prior to editing. This step helped prevent programming error.
- Reviewed items with a high nonresponse rate to catch reserve codes that were inconsistently set.

5.4 Second Follow-up File Preparation and Item Documentation

Item documentation procedures were developed to capture variable and value labels for each item. Item wording for each question was also provided as part of the documentation. This information was loaded into a documentation database that could export final data file layouts and format statements used to produce formatted frequencies for review. The documentation database also had tools to produce final ECB/input files.

Maintaining data security is a requirement that pervades all tasks, including, of course, data processing. Data security procedures in the data processing and preparation phase of the second follow-up are discussed in conjunction with the related topic of confidentiality protections associated with treatment of the analytic data (see chapter 6, section 6.6).

Chapter 6
Weighting, Imputation, and Design Effects

6.1 Overview of Weighting, Imputation, and Design Effects

Implicitly building on the sample design discussion in chapter 3, chapter 6 describes Education Longitudinal Study of 2002 (ELS:2002) weighting, imputation, and design effects for the base-year and first and second follow-up. A brief description of these three topics is provided for the base year and first follow-up; more detailed information is available from the base-year data file user's manual (NCES 2004-405) and base-year to first follow-up data file documentation (NCES 2006-344). A fuller discussion is provided for the second follow-up round (2006) of the study.

The general purpose of the ELS:2002 *weighting* scheme was to compensate for unequal probabilities of selection and to adjust for the fact that not all individuals selected into the sample actually participated. Chapter 6 sketches the school and individual sample member weights developed for the base year through second follow-up, and documents the statistical properties of the weights. *Imputation* attempts to address the issue of item nonresponse by providing a procedure that uses available information and some assumptions to derive substitute values for the missing values in a data file. The chapter provides further information on the key items that were subject to imputation, the imputation procedures, and the results of imputation. The *design effect* is a measure of sample efficiency. More specifically, the design effect is the ratio of the true variance of a statistic (taking the complex sample design into account) to the variance of the statistics for a simple random sample with the same number of cases. The chapter reports overall design effects. Since no single design effect is universally applicable to any given survey or analysis, it also reports design effects for different subgroups and statistics.

6.2 Base-Year and First Follow-up Weighting, Imputation, and Design Effects

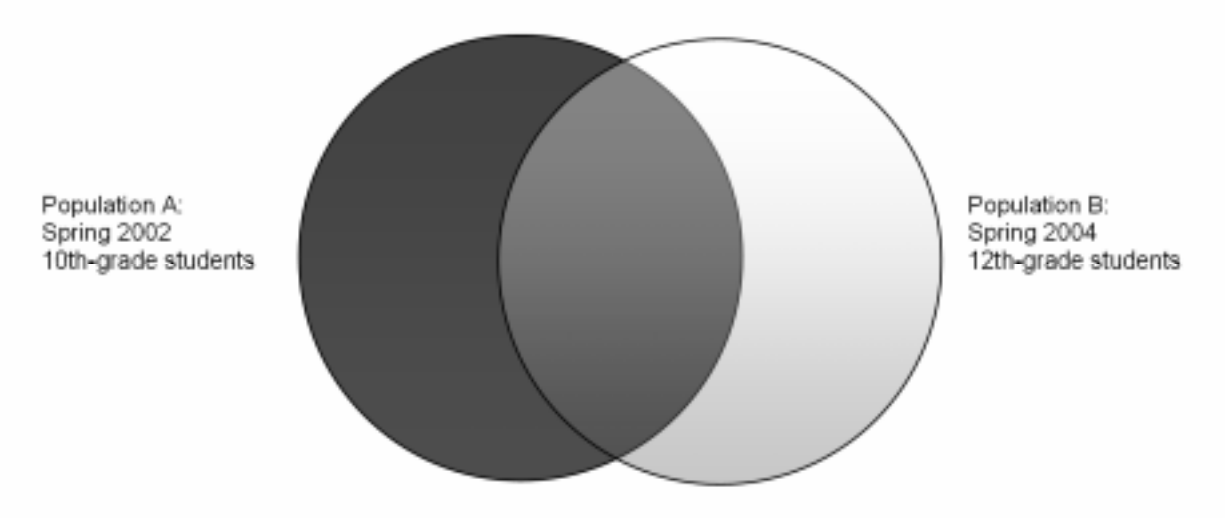
6.2.1 Calculation of Base-Year and First Follow-up Weights; Results of Weighting

6.2.1.1 Analysis Populations

The sample design for ELS:2002 supports a number of analyses, which in turn permit accurate inferences to be made to three major groups or target populations: (1) Population A: spring 2002 high school sophomores; (2) Population B: spring 2004 high school seniors; and (3) Population C: spring 2002 10th-grade schools.

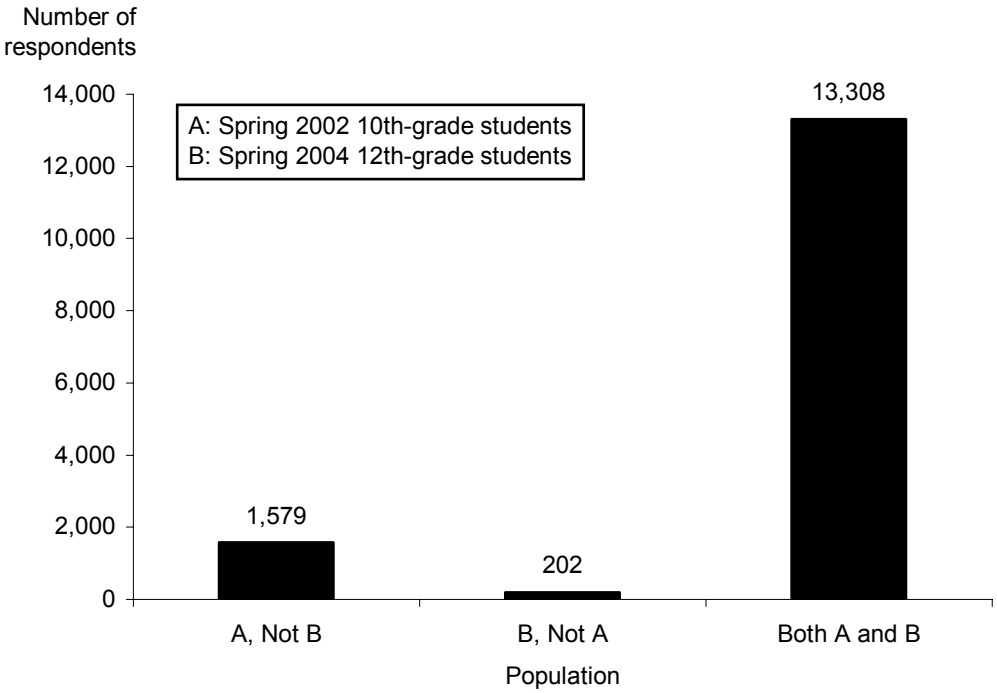
Figure 7 illustrates that whereas some students are in only population A or population B, many students are in both populations—that is, both a spring 2002 sophomore and a spring 2004 12th-grade student. Figure 8 further illustrates the overlap between the two populations.

Figure 7. Student analysis populations, by year: 2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

Figure 8. Student analysis population respondent counts, by year: 2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), and "First Follow-up, 2004."

6.2.2 Uses of Student-level Data; Student Weights

6.2.2.1 Population A: Spring 2002 Sophomores

This population can be employed in both cross-sectional and longitudinal analyses. Weights for cross-sectional analyses were created in the base year. BYSTUWT can be used for

cross-cohort comparisons of students capable of completing the questionnaire (on a cross-cohort time-lag basis employing the sophomore classes of 1980 and 1990). Students who were (by virtue of disability or language barrier) unable to complete a questionnaire were nevertheless retained in the ELS:2002 sample (and contextual data and transcripts were gathered). BYEXPWT generalizes to the entire population, including both students capable and incapable of completing the questionnaire.

The weight F1PNLWT was created for all persons who completed a questionnaire or a sufficient portion of a questionnaire, both in the base year and the first follow-up. Also, base-year data were imputed when not available from the new participant supplement (NPS) for first follow-up respondents, and these cases also have F1PNLWT. The panel weight can be used for both intracohort (across rounds of ELS:2002) and cross-cohort (longitudinal comparative analysis) purposes. An example of using a panel weight for intracohort analysis is to take a cohort of sophomores, look at their enrollment 2 years later, and determine what proportion have dropped out. An example of using a panel weight for cross-cohort analysis is to compute math gains between sophomore and senior years using the ELS:2002 panel weight and also for the National Education Longitudinal Study of 1988 (NELS:88) panel weight and then compare the gain between sophomore and senior year for the two cohorts. Missing test data were imputed, so a version of the panel weight adjusted for test nonresponse was unnecessary. The weight F1XPNLWT was created for the expanded sample of students capable and not capable of completing the questionnaire.

Base-year nonrespondents who responded in the first follow-up are considered to be part of this population, but there is no base-year weight (BYSTUWT or BYEXPWT) for them. The NPS ensured that the standard classification variables collected in the base year were also available for this group. Key variables were imputed for base-year nonrespondents who were first follow-up respondents, so that these students could be analyzed as part of the sophomore panel using F1PNLWT and/or F1XPNLWT. BYSTUWT and BYEXPWT were not recomputed.

Transcripts provide continuous data covering grades 9 through 12 for students who remained in school and were in the modal grade sequence (or a lesser range of data for students who dropped out or fell behind the modal progression). A cross-sectional 2004 transcript weight (F1TRSCWT) was produced, encompassing cases that met the following conditions for sample members for whom a transcript has been obtained: a member of the 10th- or 12th-grade cohort who was a student questionnaire completer in the base year, first follow-up, or both; or a member of the questionnaire-incapable[62] expanded sample. This weight generalizes to the analysis population of spring 2002 sophomores by subsetting the sample through the use of a flag (G10COHRT), or 2004 seniors by invoking the senior cohort flag (G12COHRT[63]).

6.2.2.2 Population B: Spring 2004 12th-Grade Students

This population can also be employed in both cross-sectional and longitudinal analyses. Weights for cross-sectional (including cross-cohort) analyses (F1QWT) were created for students capable of completing the questionnaire. This weight should be used in conjunction with a flag

[62] Questionnaire-incapable sample members were unable, owing to severe disability or language barrier, to validly be assessed or complete a student questionnaire. Nevertheless, they were not excluded from the sample. Transcripts and contextual data were collected for this group.

[63] The G12COHRT flag was updated as part of second follow-up data process activities to determine spring 2004 senior cohort membership for first follow-up nonparticipants based on responses in the 2006 data collection or transcript information.

(G12COHRT) that identifies the sample member as part of the senior cohort.[64] F1EXPWT generalizes to the entire population, including students capable and incapable of completing the questionnaire.

Note that generalizations about the mathematics achievement of the 2004 senior class involve imputation for the transfer students and other seniors who were not tested.

The cross-sectional transcript weight described also generalizes to the analysis population of spring 2004 12th-graders by subsetting the sample through the use of a flag (G12COHRT), or to the 2003–04 graduating class through the high school exit status variable, F1RTROUT.

6.2.3 Population C: Uses of School-level Data; School-level Weights

The ELS:2002 dataset supports school-level analysis using its sample of spring 2002 10th-grade schools. Weights for cross-sectional analyses were created in the base year. BYSCHWT can be used for spring 2002 10th-grade schools. In addition to the school-level data released in the base year, a restricted-use course offerings file was issued in 2006, based on course catalogues collected in the first follow-up high school transcript component.

Although it is not possible to produce a cross-sectional 2004 school weight because the first follow-up school sample is not nationally representative of American high schools in 2004, the base-year school weight can be used for longitudinal analyses treating the base-year schools as a 2002–04 panel. Although there are two data points for analysis, the weight is generalizable only to schools in 2002.

The first follow-up school data can also be analyzed using the student weight, when school data are employed as contextual information attached to the student record. That is, the school-level data (administrator questionnaire, library/media center questionnaire, facilities checklist, course offerings, school geocodes, and external data linkages) can be analyzed in relation to the sophomore or senior cohorts with the student as the primary unit of analysis. To facilitate such analyses, school-level data were replicated at the student level in the data files.

6.2.4 Base-Year and First Follow-up Weights and Their Properties

Three sets of weights were computed in the base year:

1. A school weight.
2. A weight for student questionnaire completion.
3. A contextual data weight for the "expanded" sample of both questionnaire-incapable and questionnaire-capable students (reflecting the fact that some sample members were deemed incapable of completing survey instruments owing to disability or language barriers).

Five sets of weights were computed in the first follow-up:

1. A cross-sectional weight for the expanded sample that includes sample members who completed all or a sufficient portion of the questionnaire in the first follow-up, base-year students who were still incapable of completing the questionnaire 2 years later,

[64] Note that there is a special case of the senior cohort as well: the subset of senior cohort members who in fact graduated in 2004, as contrasted to the small number of their peers who failed to graduate in their 2004 senior year.

base-year students who were newly incapable of completing the questionnaire, and freshened students who were incapable of completing the questionnaire (F1EXPWT).

2. A cross-sectional first follow-up weight for sample members who completed all or a sufficient portion of the questionnaire in the first follow-up (F1QWT).
3. A first follow-up panel weight (longitudinal weight) for the expanded sample that includes students who fully or partially completed a questionnaire in both the base year and first follow-up, students who fully or partially completed a questionnaire in the first follow-up and had base-year data imputed if not on the NPS, and students who were questionnaire incapable in the base year and/or the first follow-up (F1XPNLWT).
4. A first follow-up panel weight for sample members who fully or partially completed a questionnaire in both the base year and first follow-up or who fully or partially completed a questionnaire in the first follow-up and had base-year data imputed if not on the NPS (F1PNLWT).
5. A first follow-up weight for sample members who fully or partially participated in the transcript component was also generated (F1TRSCWT).

Additionally, there are two flags that can be used in analyses to identify members of the sophomore and senior cohorts:

1. a flag indicating a member of the sophomore cohort, that is, spring 2002 10th-grader (G10COHRT); and
2. a flag indicating a member of the senior cohort, that is, spring 2004 12th-grader (G12COHRT).

Finally, for the transcript component, a variable indicates final student status (i.e., mode of high school exit):

- A status variable that indicates whether a student is a fall 2003–summer 2004 graduate, dropout, etc. (F1RTROUT).

Table 52 through 56 show the statistical properties of the base-year and first follow-up weights.

Table 52. Statistical properties of school weight: 2002

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

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

Table 53. Statistical properties of student cross-sectional weights: 2002

Weight	BYSTUWT	BYEXPWT
Mean	223.90	223.77
Variance	18,597.52	22,448.02
Standard deviation	136.37	149.83
Coefficient of variation (x 100)	67.02	66.96
Minimum	5.09	5.09
Maximum	978.38	978.38
Skewness	0.99	0.99
Kurtosis	0.99	1.02
Sum	3,439,489.61	3,474,052.78
Number of cases	15,362	15,525

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

Table 54. Statistical properties of cross-sectional weights: 2004

Weight	F1QWT	F1EXPWT
Mean	232.29	232.36
Variance	26,283.59	26,249.80
Standard deviation	162.12	162.02
Coefficient of variation (x 100)	69.79	69.73
Minimum	1.77	1.77
Maximum	1,427.47	1,427.47
Skewness	1.21	1.21
Kurtosis	2.41	2.41
Sum	3,481,853.86	3,506,024.17
Number of cases	14,989	15,089

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

Table 55. Statistical properties of panel weights: 2004

Weight	F1PNLWT	F1XPNLWT
Mean	231.31	231.20
Variance	25,985.12	25,883.66
Standard deviation	161.20	160.88
Coefficient of variation (x 100)	69.69	69.59
Minimum	1.75	1.75
Maximum	1,445.49	1,445.49
Skewness	1.21	1.21
Kurtosis	2.48	2.49
Sum	3,403,321.11	3,441,475.79
Number of cases	14,713	14,885

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

Table 56. Statistical properties of the student transcript weight: 2004–05

Weight	F1TRSCWT
Mean	236.15
Variance	26,035.60
Standard deviation	161.36
Coefficient of variation (x 100)	68.33
Minimum	5.20
Maximum	1,125.73
Skewness	0.98
Kurtosis	0.82
Sum	3,523,285.00
Number of cases	14,920

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "High School Transcript Study, 2004–05."

Table 57 shows the interrelationships of some of the weights and flags relative to various analytic purposes.

Table 57. Relationship among weights, universe flags, populations, and respondents: 2004

Weight[1]	Universe flag	Population	Respondent
BYSTUWT	G10COHRT	Spring 2002 sophomores	Fully or partially completed questionnaire in 2002.
BYEXPWT	G10COHRT	Spring 2002 sophomores	Fully or partially completed questionnaire in 2002 or incapable of completing a questionnaire.
F1PNLWT	G10COHRT	Spring 2002 sophomores	Fully or partially completed questionnaire in 2002 and 2004 (base-year data may be imputed).
F1XPNLWT	G10COHRT	Spring 2002 sophomores	Fully or partially completed questionnaire in 2002 and 2004 (base-year data may be imputed) or incapable of completing a questionnaire in 2002 or 2004.
F1QWT	G10COHRT	Spring 2002 sophomores	Fully or partially completed questionnaire in 2004.
	G12COHRT	Spring 2004 seniors	
F1EXPWT	G10COHRT	Spring 2002 sophomores	Fully or partially completed questionnaire in 2004 or incapable of completing a questionnaire in 2004.
	G12COHRT	Spring 2004 seniors	
F1TRSCWT	G10COHRT	Spring 2002 sophomores	Fully or partially completed student transcript data.
	G12COHRT	Spring 2004 seniors	
	F1RTROUT	High school graduating class of 2004	

[1] The expanded sample weights and the full expanded sample are available on the restricted-use file but not on the public-use file.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

6.2.5 Base-Year and First Follow-up Item Imputation

The imputation procedures used for the base-year and first follow-up study include logical imputation, a weighted sequential hot deck procedure, and a multiple imputation procedure. Eighteen variables were selected for imputation. Four were unique to the first follow-up, and 14 were key demographic and family background variables that were chosen for imputation in the base year and first follow-up. These key variables were imputed when not provided by respondents in the base-year questionnaire or the first follow-up new participant supplement for first follow-up respondents. In the first follow-up, missing key variables were imputed for sample members who were one of the following: base-year nonrespondents, 12th-grade spring-term freshened sample members, or base-year questionnaire-incapable students (who were part of the base-year expanded sample only). Additionally, the 10th-grade student ability estimates for mathematics and reading were imputed for the base-year nonrespondents who became first follow-up respondents since they were included in the spring 2002 sophomore cohort. These ability estimates had been imputed, if missing, in the base year for base-year respondents.

Two first follow-up variables were imputed, as applicable, when the data were missing. Student enrollment status as of spring 2004 was imputed for the first follow-up respondents if enrollment status was not provided by the sample school. The first follow-up mathematics ability estimate was imputed, if missing, for first follow-up respondents who were considered in-school students: students at the base-year school or at another (transfer) school as of spring 2004. (Sample members who dropped out, finished high school early, or were being homeschooled as of spring 2004 were not defined as in-school students, so no ability estimates were determined

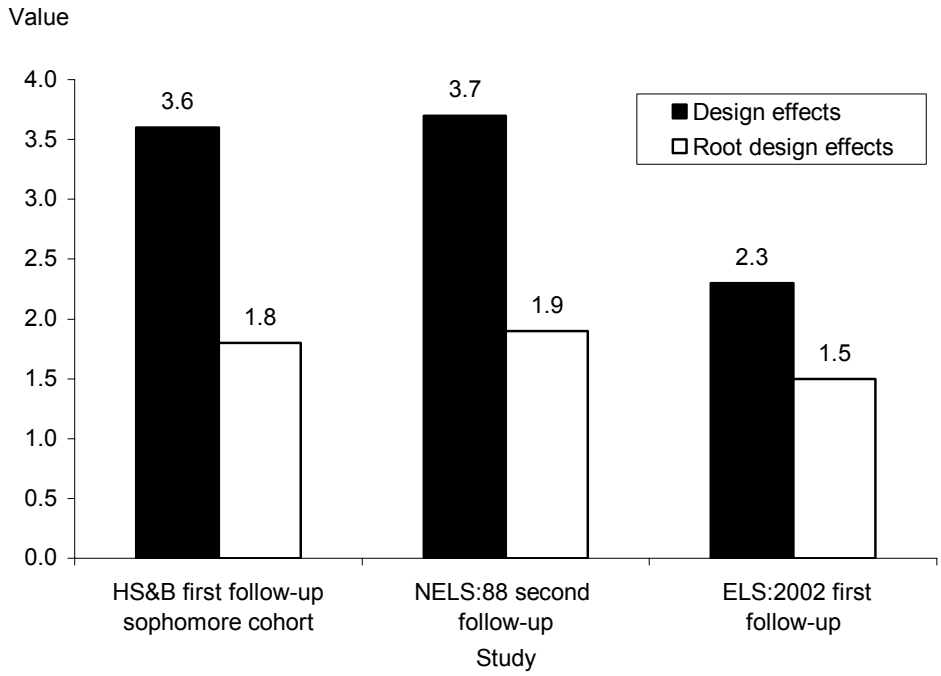
for them.) Only students still at the base-year schools were tested—ability estimates were imputed for all transfer student respondents.

6.2.6 Base-Year and First Follow-up Standard Errors and Design Effects

The variance estimation procedure had to take into account the ELS:2002 complex sample design, including stratification and clustering. One common procedure for estimating variances of survey statistics is the Taylor series linearization procedure. This procedure takes the first-order Taylor series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. For stratified multistage surveys, the Taylor series procedure requires analysis strata and analysis primary sampling units (PSUs) (in ELS:2002, schools are the PSUs). Therefore, analysis strata and analysis PSUs were created in the base year and used again in the first follow-up. The impact of the departures of the ELS:2002 complex sample design from a simple random sample design on the precision of sample estimates can be measured by the design effect, as reported in this document and the prior manuals. Taylor series estimation was used for the base year and first follow-up. ELS:2002 base-year and first follow-up data are available as public- or restricted-use electronic codebook (ECB) systems. The data are also available in a Data Analysis System (DAS). For the DAS, balanced repeated replication (BRR) replicate weights are used.

Figure 9 shows ELS:2002 design effects in historical perspective, that is, displayed in comparison to design effects in NELS:88 and High School and Beyond (HS&B). These have been calculated on the full sample (i.e., for NELS:88 and ELS:2002, all cohorts combined).

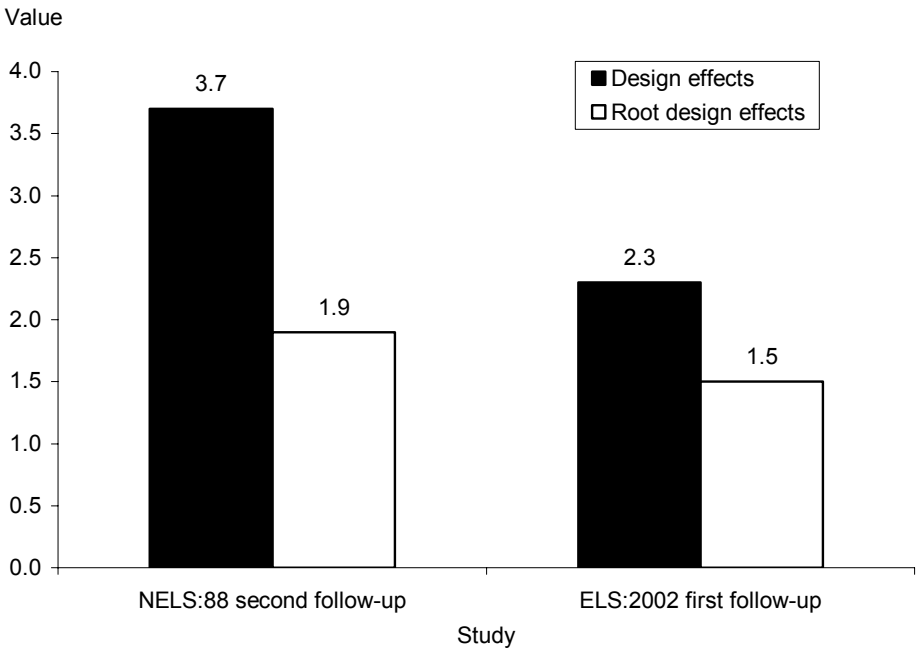
Figure 9. Full sample mean design effects and root design effects, by longitudinal study: Selected years, 1982–2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, High School and Beyond (HS&B), "First Follow-up, 1980"; National Education Longitudinal Study of 1988 (NELS:88), "Second Follow-up, 1992"; and Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

The design effects indicate that the ELS:2002 panel sample (sophomore cohort) was more efficient than the NELS:88 first and second follow-up (F1F2) panel sample (sophomore cohort). For means and proportions based on first follow-up questionnaire data for all respondents, the average design effect in ELS:2002 was 2.23; the comparable figure was 3.73 for the NELS:88 sophomore cohort. Figure 10 shows the mean design effects and root design effects for the NELS:88 second follow-up and the ELS:2002 first follow-up sophomore cohort. The difference in design effects is also apparent for some subgroup estimates. Ingels et al. (1994a) present design effects for 16 subgroups defined similarly to those used in the ELS:2002 analysis (Ingels et al. 2005b, table 25). For all 16 subgroups, the ELS:2002 design effects are smaller on average than those for the NELS:88 sophomore cohort.

Figure 10. Mean design effects and root design effects, by NELS:88 and ELS:2002 panel sample (sophomore cohort): 1992 and 2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), "Second Follow-up, 1992"; and Education Longitudinal Study of 2002 (ELS:2002), "First Follow-up, 2004."

The design effects indicate that the ELS:2002 full and panel samples were also more efficient than the NELS:88 sample for dropouts. For means and proportions based on first follow-up questionnaire data for dropouts, the average design effect in ELS:2002 was 1.31 for both the full and panel samples; the comparable figures were 2.9 and 2.8 for the NELS:88 second follow-up full and F1F2 panel samples, respectively.

The smaller design effects in ELS:2002 compared with those for NELS:88 and HS&B are probably due to the higher rates of subsampling in the latter two studies. Additionally, disproportional strata representation was introduced in the NELS:88 first follow-up, when students dispersing between 8th and 10th grade were severely subsampled. See Ingels et al. (1994b) for more details. In HS&B, the sophomore cohort members who were no longer in the base-year school were subsampled. See Spencer, Sebring, and Campbell (1987) for more details. The general tendency in longitudinal studies is for design effects to lessen over time, as

dispersion reduces the original clustering. Subsampling increases design effects because it introduces additional variability into the weights with an attendant loss in sample efficiency.

The smaller design effects in ELS:2002 compared with those for the HS&B sophomore cohort also may reflect the somewhat smaller student cluster size used in the later survey in the base year. Although the clusters were reduced somewhat in the first follow-up for both studies, a number of students remained in the base-year school. The HS&B base-year sample design called for 36 sophomores selected from each school. The ELS:2002 sample design called for about 26 sophomores selected from each school.

6.2.7 First Follow-up Transcript Component Design Effects

Within the transcript component, standard errors and design effects were computed for the entire sample and for the following subgroups:

- sex (male and female);
- race/ethnicity (American Indian or Alaska Native, Asian or Pacific Islander, Black or African American, Hispanic or Latino, More than one race, and White and all other races);[65]
- school sector (public, Catholic, and other private);
- socioeconomic status (SES) (lowest quarter, middle two quarters, and highest quarter); and
- school urbanicity (urban, suburban, and rural).

Additionally, standard errors and design effects were computed for spring 2004 graduates with complete transcript information and for the above subgroups. Table 58 summarizes the average transcript mean design effects (DEFFs) and root design effects (DEFTs) for the full sample for all respondents and each subgroup. Table 59 summarizes the average transcript DEFFs and DEFTs for the spring 2004 graduates with complete transcript information for all respondents and each subgroup. Appendix G contains tables of transcript design effects for specific variables for different subpopulations.[66] The standard errors and design effects were calculated using the transcript weight (F1TRSCWT). Each table includes the survey item (or composite variable), variable name and value for categorical variables, percent estimate, design standard error, simple random sample standard error, sample size (N), DEFF, and DEFT. Note that the mean DEFTs reported in this table were not calculated directly from the mean DEFF but, rather, were the average of the DEFTs over the items shown in each table in appendix G. Therefore, readers cannot derive the DEFT using the DEFF reported in table 58 and table 59. See section 3.5.2 of Ingels et al. (2005b) for more details about design effects.

[65] "White and all other races" is predominantly White, with a very small number of individuals from other race categories. All race categories exclude individuals of Hispanic or Latino origin.

[66] It is important to compare design effects across cohorts (e.g., ELS:2002 versus NELS:88), so table 3.3-1 from the *NELS:88 Second Follow-Up: Transcript Component Data File User's Manual* (Ingels et al. 1995) was initially used to help guide the selection of variables. However, the ELS:2002 variables chosen differ somewhat from those used in constructing design effects for NELS:88 because there were considerable differences in the types and composition of variables produced in each study. Nonetheless, the variables presented are a good representation of the breadth of information obtained from the transcripts. These items should provide a range of design effects that will give a reasonable average for both the entire sample and for analytically important subgroups.

The BY-F1 DFD (Ingels et al. 2005b) shows in its table 25 the design effects for the first follow-up full sample. With the exception of respondents who reported more than one race, the design effects are higher in the transcript study than in the first follow-up. For example, of the 30 variables used to compute design effects, the mean is 4.56 for all transcript respondents and 2.26 for all first follow-up respondents.

Table 58. Mean design effect and root design effect for the ELS:2002 high school transcript study, by selected student characteristics: 2004–05

Characteristic	Mean design effect	Mean root design effect
All respondents	4.57	2.12
Sex		
Male	2.95	1.71
Female	3.32	1.81
Race/ethnicity[1]		
American Indian or Alaska Native	1.69	1.28
Asian or Pacific Islander	2.68	1.63
Black or African American	2.24	1.48
Hispanic or Latino	3.04	1.73
More than one race	1.70	1.30
White and all other races	3.51	1.85
School sector		
Public	4.00	1.98
Catholic	7.00	2.54
Other private	7.92	2.76
Socioeconomic status (SES)		
Lowest quarter	2.34	1.52
Middle two quarters	2.93	1.70
Highest quarter	2.85	1.67
Urbanicity		
Urban	5.90	2.41
Suburban	3.93	1.96
Rural	4.17	2.00

[1] "White and all other races" is predominantly White, with a very small number of individuals from other race categories. All race categories exclude individuals of Hispanic or Latino origin.

NOTE: The mean root design effect was not calculated directly from the mean design effect but, rather, is the average root design effect over selected items.

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

Table 59. Mean design effect and root design effect for 2004 high school graduates from the ELS:2002 transcript data, by selected student characteristics: 2004–05

Characteristics	Mean design effect	Mean root design effect
2004 high school graduates	4.29	2.04
Sex		
Male	2.84	1.67
Female	3.25	1.79
Race/ethnicity[1]		
American Indian or Alaska Native	1.88	1.34
Asian or Pacific Islander	2.61	1.61
Black or African American	2.26	1.49
Hispanic or Latino	2.59	1.59
More than one race	1.86	1.36
White and all other races	3.44	1.81
School sector		
Public	3.69	1.89
Catholic	7.41	2.63
Other private	7.43	2.66
Socioeconomic status (SES)		
Lowest quarter	2.07	1.43
Middle two quarters	2.85	1.67
Highest quarter	2.79	1.65
Urbanicity		
Urban	5.37	2.30
Suburban	3.83	1.93
Rural	4.09	1.95

[1] "White and all other races" is predominantly White, with a very small number of individuals from other race categories. All race categories exclude individuals of Hispanic or Latino origin.

NOTE: The mean root design effect was not calculated directly from the mean design effect but, rather, is the average root design effect over selected items.

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

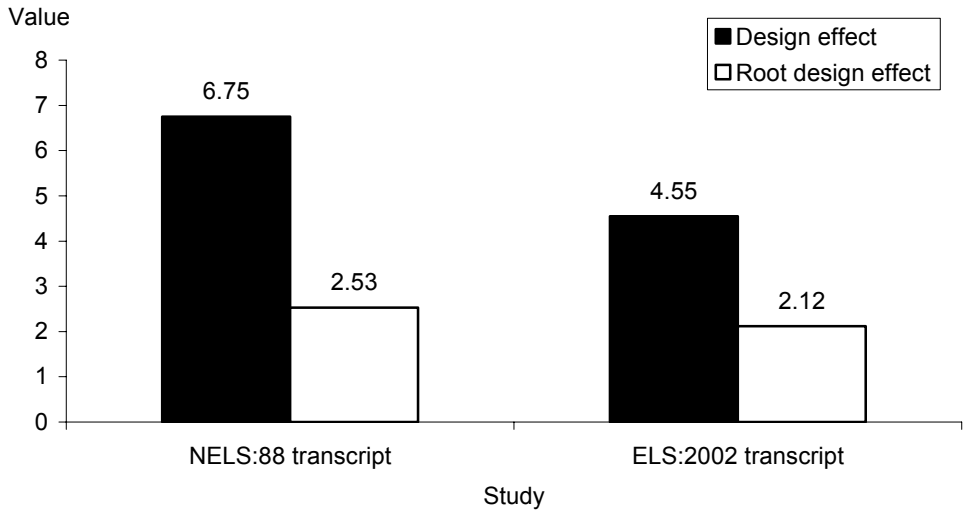
The magnitude of design effects is affected mainly by the degree of clustering in the sample and the variability of the analysis weights. The degree of clustering is determined by the cluster size and the intraclass correlation. The cluster size is the number of respondents from each school. The intraclass correlation is the correlation between the data points for any two students selected from a given school. The variability of the transcript weights can be measured by computing the unequal weighting effect (UWE). The overall UWE is slightly smaller for the transcript study than for the first follow-up (1.47 compared with 1.49), so the variability of the weights is probably not much of a cause for the difference in the design effects between the transcript study and the first follow-up.

Many of the large transcript design effects are for the variables indicating Carnegie units (CUs) earned in certain subject areas. For example, the variable "Total CUs in social studies" frequently has one of the highest design effects, because these units are nearly identical for all

students in many schools. The largest design effect is "Total CUs in general labor market preparation" for Catholic school students, because such students typically have zero CUs in this subject area.

Figure 11 shows the DEFFs and DEFTs for both the ELS:2002 and NELS:88 high school transcript studies. The design effects indicate that the ELS:2002 high school transcript sample was more efficient than the NELS:88 high school transcript sample. The average design effect in ELS:2002 was 4.57; the comparable figure was 6.75 for NELS:88.[67]

Figure 11. Mean design effect and root design effect for the ELS:2002 and NELS:88 high school transcript studies: 1992–93 and 2004–05



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), "High School Transcript Study"; and Education Longitudinal Study of 2002 (ELS:2002), "High School Transcript Study."

6.2.8 Base-Year and First Follow-up Disclosure Risk Analysis and Protection

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

[67] The difference in design effects is also apparent for some subgroup estimates. In NELS:88, design effects were produced for 18 subgroups, 16 of which are defined similarly to those in ELS:2002. For 15 of the 16 subgroups, the ELS:2002 design effects are smaller on average than those for NELS:88. These smaller design effects in ELS:2002 compared with those for NELS:88 are probably due to the magnitude of subsampling in the first follow-up (1990) of NELS:88. High school transcript component design effects from NELS:88 can be found in Ingels et al. (1995).

6.2.9 Base-Year and First Follow-up Nonresponse Bias Analyses

The overall weighted school response rate was 68 percent in the base year. A follow-up survey of nonresponding schools was used to collect basic school characteristics needed to support comparisons with the participating schools. Some 93 percent of the nonparticipating schools responded to the nonresponse follow-up survey. Some sample frame data were also available for both responding and nonresponding schools. Results of the analysis showed only a small potential for bias. The identified variables were used to inform nonresponse adjustments for the base-year schools and students. Bias due to nonresponse was estimated both prior to computing weights and after computing weights. For details see Ingels et al. (2004).

The overall weighted student response rate was 87 percent in the base year (2002). Overall response was 89 percent (including all groups, e.g., students, transfer students, dropouts) in the first follow-up (2004). Student unit nonresponse bias analyses were performed in both the base year and first follow-up. An item nonresponse bias analysis was also performed for all questionnaire variables in which response fell below 85 percent. Details of the bias analyses are given in Ingels et al. (2004, 2005b). See also appendix H of this document.

6.3 Calculation of Second Follow-up Weights and Results of Weighting

A variety of topics are discussed in the following subsections. Sections 6.3.1 and 6.3.2 provide a high-level overview of the ELS:2002 target populations and potential domains of analysis for those populations and describe the analysis weights created for the second follow-up. Section 6.3.2 also lists the names of the analysis weights created for the second follow-up and lists the names of the flags used to restrict analyses to the target populations of the ELS:2002 study.

The model-based approached for weight adjustment is discussed in section 6.3.3.[68] The list of variables used in the nonresponse models is also provided in section 6.3.3. The Chi-squared automatic interaction detection analysis (CHAID) used to identify interaction terms included in the nonresponse models is described in section 6.3.3.

Details of the weight adjustment factors used to create the second follow-up analysis weights are given in sections 6.3.4, 6.3.5, and 6.3.6. A discussion of the BRR weights produced for the second follow-up DAS occurs in section 6.3.7 and a brief discussion of quality control methods used to produce the second follow-up weights may be found in section 6.3.8. BRR weights are also included with the ECB.

6.3.1 Target Populations and Analysis Domains

The sample design for ELS:2002 was developed so that relevant samples, suitably weighted, would be representative of three target populations: spring-term 2002 10th-grade students, spring-term 2004 12th-grade students, and spring-term 2002 10th-grade schools.

[68] *Propensity modeling* approaches were used in nonresponse adjustment for the NELS:88 school, and ELS:2002 school and student weights. For the NELS:88 student weights, and school and student weights in HS&B and NLS:72, a *weighting cell* approach was used. A comparison of the two approaches is included in appendix H of this volume. The comparison shows that the two methods generate very similar results, and so are unlikely to be a source of noncomparability between ELS:2002 and the prior studies.

Within these three target populations are a variety of important analysis domains. These analysis domains are subsets of the three target populations and, while these subsets are themselves populations, the ELS:2002 sample design does not guarantee that the ELS:2002 sample will be representative of all subsets of the three primary target populations. The following lists give examples of analytic domains as subsets of the three target populations.

Population A: Spring-term 2002 10th-grade students:

- Domains[69]
  - Spring 2002 10th-grade students capable of completing the student questionnaire;
  - Spring 2002 10th-grade students in base-year school in spring 2004;
  - Spring 2002 10th-grade students in a different school in spring 2004 (transfers);
  - Spring 2002 10th-grade students who were dropouts in spring 2004;
  - Spring 2002 10th-grade students who graduated or achieved equivalency early (i.e., on or before March 15, 2004);
  - Spring 2002 10th-grade students who graduated by August 31, 2004;
  - Spring 2002 10th-grade students who were homeschooled in spring 2004;[70]
  - Spring 2002 White 10th-grade students;
  - Spring 2002 Black 10th-grade students;
  - Spring 2002 Hispanic 10th-grade students;
  - Spring 2002 Asian 10th-grade students;
  - Spring 2002 public school 10th-grade students; and
  - Spring 2002 private school 10th-grade students.

Population B: Spring-term 2004 12th-grade students:

- Domains
  - Spring 2004 12th-grade students capable of completing the student questionnaire;
  - Spring 2004 12th-grade students regardless of final spring 2004 graduation status;
  - Spring 2004 12th-grade students who graduated by August 31, 2004;
  - Spring 2004 White 12th-grade students;
  - Spring 2004 Black 12th-grade students;

[69] The domains listed are important domains but are not the only possible domains.

[70] Although conceptually spring 2002 sophomores who were homeschooled in 2004 may be thought of as an analysis population, they were not designed to be so and were therefore not subject to minimum sample size requirements. The group is of limited analytic utility owing both to the low sample size and to the narrowness of the population definition. The compelling practical reason for distinguishing this group was so that they could be administered only those items consonant with their unique situation as out-of-school students.

  - Spring 2004 Hispanic 12th-grade students;
  - Spring 2004 Asian 12th-grade students;
  - Spring 2004 public school 12th-grade students; and
  - Spring 2004 private school 12th-grade students.

Population C: Spring 2002 10th-grade schools:

- Domains
  - School type: public, Catholic, and other private;
  - Urbanicity: urban, suburban, and rural;[71] and
  - Region: Northeast, Midwest, South, West.

ELS:2002 student sample members were interviewed as part of second follow-up activities. Sample members who completed a certain prespecified proportion of the second follow-up questionnaire were considered to be second follow-up respondents. ELS:2002 second follow-up respondents may be in either population A (10th-grade cohort), or population B (12th-grade cohort), or in both. In order to identify those respondents belonging to a particular target population, two flag variables are provided. The flag G10COHRT denotes membership in the spring 2002 10th-grade population and the flag G12COHRT[72] denotes membership in the spring 2004 12th-grade population. Figure 12 shows the distribution of ELS:2002 second follow-up respondents with respect to the two student target populations.

Analytic uses of these three populations, and the weighting required to support the analyses, are discussed in section 6.3.2.

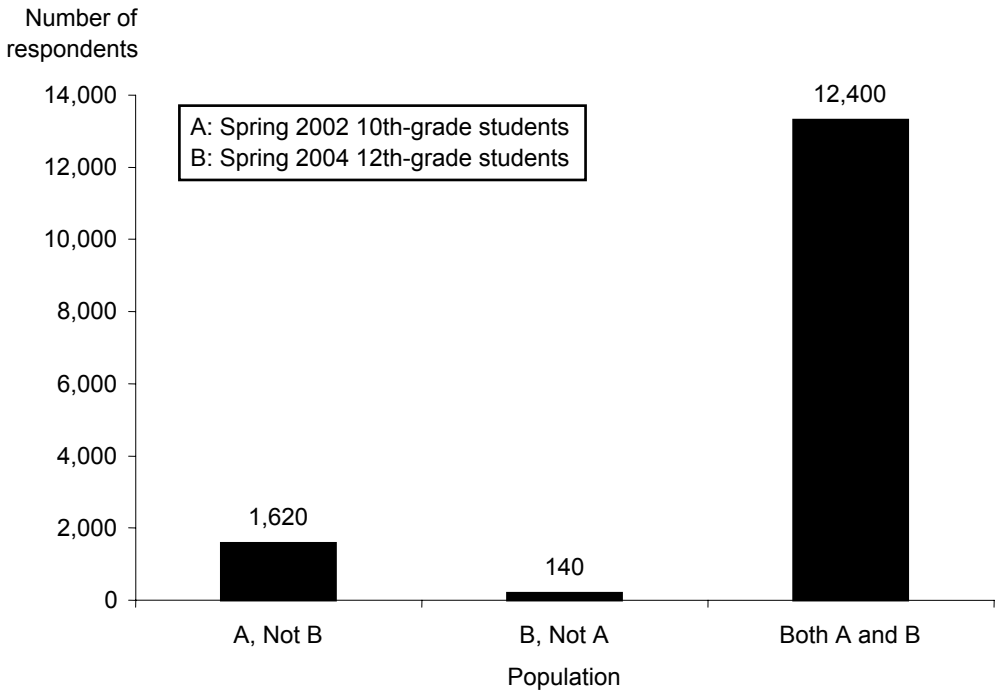
6.3.2 Overview of Second Follow-up Analysis Weights

The analysis weights for the ELS:2002 second follow-up were created in order to allow for analysis of the spring 2002 10th-grade population and the spring 2004 12th-grade population. Since the ELS:2002 study is longitudinal, analyses of these two populations may focus on characteristics of these populations at one point in time or may focus on how characteristics of these populations vary over time. Second follow-up cross-sectional weights were created to allow for analysis of these two populations in 2006 and panel weights were created to allow for analysis of these two populations over multiple rounds of the ELS:2002 study.

[71] NCES has recently changed its locale code system. The new codes draw on a four-part classification: city, suburban, town, and rural. Cities and suburbs are further divided into small, mid-size, and large, and towns and rural areas can be related (via measures of proximity) to urbanized areas (urban fringe, distant, remote). While the tripartite classification was used in ELS:2002 sampling, any analyst who wants to employ the new locale codes with the ELS:2002 base-year and transfer schools can use the ELS:2002 links to the Common Core of Data (CCD) and Private School Survey (PSS) databases to do so.

[72] G12COHRT includes members of the senior cohort determined in the first follow-up (G12COHRT = 1) as well as those whose membership status was determined in the second follow-up (G12COHRT = 2).

Figure 12. Student analysis population respondent counts, by cohort: 2006



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Four sets of weights were computed for the second follow-up:

- A cross-sectional weight for sample members who responded[73] in the second follow-up (F2QWT).
- A cross-sectional transcript weight for sample members who responded in the second follow-up and for whom a transcript was collected in the first follow-up transcript component (F2QTSCWT).
- A second follow-up panel weight (longitudinal weight) for all sample members who responded in the second follow-up and responded in the first follow-up (F2F1WT).
- A second follow-up panel weight for all sample members who responded in the second follow-up and responded in the base year (F2BYWT), or who were base-year nonrespondents but for whom the base-year classification variables were collected in the first follow-up and their base-year test scores imputed.

These weights and the types of analyses that may be conducted using these weights are described below. While second follow-up student weights were created, no second follow-up school weights were created. Discussion of school weights in the context of the ELS:2002 second follow-up may be found in section 6.3.2.3.

[73] In the base year and first follow-up of ELS:2002, sample members are considered part of the expanded analysis sample if they complete at least a certain proportion of the round-appropriate questionnaire or if they are "questionnaire-incapable" for that round (though eligible for contextual data and transcripts). (Again, questionnaire-incapable students were those who could not be validly assessed or surveyed owing to severe disability or language barrier.) Sample members are considered respondents in the second follow-up if they complete at least a certain proportion of the second follow-up questionnaire.

6.3.2.1 Cross-sectional Weights

Two cross-sectional weights were constructed for the ELS:2002 second follow-up study. The first cross-sectional weight, F2QWT, was constructed so that the population of spring 2002 10th-grade students and the population of spring 2004 12th-grade students could be analyzed using respondent data collected in the ELS:2002 second follow-up. The second cross-sectional weight, F2QTSCWT, encompasses cases that meet the following conditions: (a) member of the 10th- or 12th-grade cohort who had a first follow-up transcript, and (b) a second follow-up respondent. This transcript weight allows for analysis of both target student populations using those second follow-up respondents who had transcript data collected in the first follow-up.

As noted in section 6.3.1, second follow-up respondents may be in the population of spring 2002 10th-graders, may be in the population of spring 2004 12th-graders, or may be in both populations. Analyses designed to assess characteristics of one of the populations must take care to restrict analyses to those second follow-up respondents in the population of interest. In order to identify those second follow-up respondents who are members of the two student populations, two flag variables, G10COHRT and G12COHRT, are provided in the restricted-use file. Those second follow-up respondents with a value of 1 for G10COHRT are members of the population of spring 2002 10th-graders. Those second follow-up respondents with a value of 1 (determined in the first follow-up or a value of 2 determined in the second follow-up) for G12COHRT are members of the population of spring-term 2004 12th-graders.

The two cross-sectional weights may be used to analyze both student populations as long as the two cohort flag variables, G10COHRT and G12COHRT, are used to select those second follow-up respondents who belong to the student population of interest. Note that if these flag variables are not used in analysis, then the set of all second follow-up respondents represents the union of the population of spring 2002 10th-graders with the population of spring 2004 12th-graders. The union of these two populations includes individuals who were in the 10th grade in spring term 2002, who were in the 12th grade in spring term 2004, or who were both in the 10th grade in spring term 2002 and in the 12th grade in spring term 2004. Such individuals may be in one or both of the target student populations.

6.3.2.2 Panel Weights

Two panel weights were constructed for the ELS:2002 second follow-up study. The purpose of creating these panel weights was to facilitate analyses designed to examine how the two student populations change over time. The panel weights can be used for both intracohort (across rounds of ELS:2002) and cross-cohort (longitudinal comparative analysis) purposes. An example of using a panel weight for intracohort analysis is to take a cohort of sophomores in 2002 and determine what proportion had enrolled in a postsecondary institution by 2006. An example of using a panel weight for cross-cohort comparison would be to model the transition from high school to postsecondary outcomes, comparing the four senior cohorts—NLS:72 (1972), HS&B (1980), NELS:88 (1992), and ELS:2002 (2004)—2 years after high school graduation.

The panel weight, F2BYWT, was produced for all ELS:2002 sample members who responded[74] in the base year and in the second follow-up, or who responded in the second follow-up and had key base-year data that were collected in the first follow-up. The set of sample members who responded in the base year and second follow-up is only representative of the population of spring 2002 10th-grade students; only sample members who are members of the 10th-grade cohort will have a nonmissing value for this panel weight. It is not necessary[75] to use the flag variable G10COHRT in conjunction with this panel weight since, by construction, only second follow-up respondents who are members of the spring 2002 10th-grade population will have a nonzero value for the panel weight.

The panel weight, F2F1WT, was produced for all sample members who responded in the first and second follow-ups. This panel weight will generalize to the population of spring 2002 10th-grade students and will generalize to the population of spring 2004 12th-grade students when used in conjunction with the two flag variables G10COHRT and G12COHRT, respectively.

As noted in the ELS:2002 base-year to first follow-up data file documentation, base-year nonrespondents who responded in the first follow-up are considered to be members of the spring 2002 10th-grade population, but there is no base-year weight (BYSTUWT or BYEXPWT) for them. The new participant supplement employed in the first follow-up ensured that the standard classification variables collected in the base year were also available for this group. However, key variables were imputed for base-year nonrespondents who were first follow-up respondents,[76] so that these students could be analyzed as part of the sophomore panel using F1PNLWT and/or F1XPNLWT. These students who are second follow-up respondents may also be analyzed as part of the sophomore panel using F2F1WT and F2BYWT.

6.3.2.3 School Weights and the Second Follow-up

The second follow-up to the ELS:2002 study surveyed base-year and first follow-up sample members but did not attempt to survey the ELS:2002 base-year schools. Since most of the ELS:2002 sample members were out of high school in 2006, the utility of information collected from base-year sampled schools as part of the second follow-up would have been extremely limited.

Although it is not possible to produce a cross-sectional 2006 school weight because the second follow-up school sample is not nationally representative of American high schools in 2006, the base-year school weight can be used for longitudinal analyses treating the base-year schools as a panel. Although there are multiple data points for analysis, the weight maintains the generalizibility only to schools in 2002.

---

[74] Sample members who did not respond in the base year but did respond in the first follow-up were given a new participant supplement questionnaire in order to gather some of the same information that was collected on base-year respondents. Consequently, these base-year nonrespondents who responded in the first follow-up were treated as base-year respondents in the construction of first follow-up panel weights. These sample members were treated as base-year respondents in the construction of second follow-up panel weights.

[75] It is possible that statistical software not designed for the analysis of sample survey data may fail to exclude records that have analysis weights of zero. The G10COHRT flag may be used to specifically restrict analyses to members of the 10th-grade cohort in order to avoid such a situation from arising.

[76] However, sample members who met the dual conditions of being (1) base-year nonrespondents and (2) questionnaire-incapable in the first follow-up were given a cross-sectional weight in the first follow-up but were not given a panel weight, nor, owing to lack of information, were base-year data imputed for them.

6.3.2.4 Second Follow-up Weights and Prior-Round Weights

In both the base year and first follow-up of the ELS:2002 study, some sample members were not able to complete the sample member questionnaires because of limited English proficiency or because of physical or mental limitations. However, information could be collected from individuals, such as school administrators, parents, and teachers associated with these sample members. In a given prior round, the set of respondents in that round combined with the set of sample members who were questionnaire-incapable was referred to as the expanded sample for that round. Expanded sample weights that encompass both the questionnaire-capable and questionnaire-incapable sample were included only in restricted-use files.

Unlike the prior rounds, any prior-round questionnaire-incapable sample member who was unable to complete the second follow-up questionnaire was considered out of scope for the second follow-up. Since all second follow-up questionnaire-incapable sample members were considered to be out of scope, no second follow-up expanded sample weights were constructed.

There are several flags that can be used in analyses to identify members of the sophomore and senior cohorts:

- a flag indicating a member of the sophomore cohort, that is, spring 2002 sophomore (G10COHRT); and
- a flag indicating a member of the senior cohort, that is, spring 2004 senior (G12COHRT).

Table 60 summarizes the ELS:2002 analysis weights and the associated universe flags, populations (described in section 6.3.1), and respondents.

Table 60. Relationship among weights, universe flags, populations, and respondents: 2002–06

Weight	Universe flag	Population	Respondent
BYSTUWT	G10COHRT	A—Spring 2002 10th-grader	Fully or partially completed questionnaire in 2002
BYEXPWT	G10COHRT	A—Spring 2002 10th-grader	Fully or partially completed questionnaire in 2002 or incapable of completing a questionnaire
F1PNLWT	G10COHRT	A—Spring 2002 10th-grader	Fully or partially completed questionnaire in 2002 and 2004 (base- year data may be from the new participant supplement or imputed)
F1XPNLWT	G10COHRT	A—Spring 2002 10th-grader	Fully or partially completed questionnaire in 2002 and 2004 (base-year data may be from the new participant supplement or imputed) or incapable of completing a questionnaire in 2002 or 2004
F1QWT	G10COHRT G12COHRT	A—Spring 2002 10th-grader B—Spring 2004 12th-grader	Fully or partially completed questionnaire in 2004
F1EXPWT	G10COHRT G12COHRT	A—Spring 2002 10th-grader B—Spring 2004 12th-grader	Fully or partially completed questionnaire in 2004 or incapable of completing a questionnaire in 2004
F1TRSCWT	G10COHRT G12COHRT	A—Spring 2002 10th-grader B—Spring 2004 12th-grader	Fully or partially completed transcript data and fully or partially completed first follow-up or base-year questionnaire or members of the expanded sample
F2QWT	G10COHRT G12COHRT	A—Spring 2002 10th-grader B—Spring 2004 12th-grader	Fully or partially completed questionnaire in 2006
F2QTSCWT	G10COHRT G12COHRT	A—Spring 2002 10th-grader B—Spring 2004 12th-grader	Fully or partially completed questionnaire in 2006 and full or partial transcript data
F2F1WT	G10COHRT G12COHRT	A—Spring 2002 10th-grader B—Spring 2004 12th-grader	Fully or partially completed questionnaire in 2004 and 2006 or incapable of completing a questionnaire in 2004 and fully or partially completed questionnaire in 2006
F2BYWT	G10COHRT	A—Spring 2002 10th-grader	Fully or partially completed questionnaire in 2002 and 2006 or incapable of completing a questionnaire in 2002 and fully or partially completed questionnaire in 2006

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002," "First Follow-up, 2004," and "Second Follow-up, 2006."

6.3.3 Overview of Nonresponse and Calibration Methodology

All second follow-up analysis weights were created by applying a variety of weight adjustments to the second follow-up base weight (discussed in section 6.3.4). These weight adjustments were designed to account for three issues:

- Some ELS:2002 sample members were not fielded for the second follow-up.

- Some of the ELS:2002 sample members fielded for the second follow-up did not respond.
- Application of weight adjustments to account for the first two issues resulted in weight sums for key analysis domains that differed from prior-round weight sums.

Two simple ratio adjustments were applied to the second follow-up base weight in order to account for the first issue and details of this adjustment are given in section 6.3.4.

The most significant and complex weight adjustments are related to the second and third issues. The weight adjustments associated with the second issue are known as nonresponse adjustments. Two types of nonresponse occurring during second follow-up data collection were considered: nonresponse arising from the inability to locate or contact a sample member and nonresponse arising from sample member refusal to participate once contacted. After examining the nonresponse cases occurring because of refusal and the nonrespondent cases occurring because of inability to locate or contact, a determination was made to treat all nonrespondents as one group.

The weight adjustments associated with the third issue are known as poststratification or calibration[77] adjustments. As the ELS:2002 second follow-up sample weights are not adjusted to sum to population totals, the adjustments associated with the third issue are referred to as calibration adjustments.

In addition to the nonresponse and calibration adjustments described above, the second follow-up transcript weight included two nonresponse adjustments followed by a subsequent calibration adjustment; the first adjustment accounted for nonresponse arising from the student's school refusing to provide a transcript, the second adjustment accounted for nonresponse resulting from student refusal to allow the transcript information to be included with the ELS:2002 data, and the third adjustment calibrated weight sums to prior-round totals.

While there are several methods[78] that may be used to adjust sampling weights to account for nonresponse and to calibrate weight sums, the method used to create the ELS:2002 second follow-up analysis weights followed a model-based approach, which is given below. Specific details of the nonresponse and calibration adjustments applied to produce the second follow-up analysis weights may be found in sections 6.3.5 and 6.3.6.

6.3.3.1 Generalized Exponential Model

All nonresponse and calibration adjustments were calculated using RTI's generalized exponential modeling procedure (GEM) (Folsom and Singh 2000), which is similar to logistic modeling with bounds for adjustment factors.

The GEM approach is a general version of weighting adjustments and was based on a generalization of Deville and Särndal's logit model (Deville and Särndal 1992). GEM is not a competing method to weighting classes or logistic regression; rather, it is a method of creating

[77] Poststratification typically refers to the process of adjusting sample weights so that the weights sum to population totals derived from sources external to the sample of interest. Calibration is used to denote adjusting weight sums to sum to prior-round totals.

[78] For example, at the school and student level in HS&B, and at the student level only in NELS:88, a weighting cell approach to nonresponse adjustment was used. For a comparison of propensity model versus weighting cell approaches, see the paper by Siegel, Copello, and Chromy that appears as appendix H of this report.

weight adjustments that provides a wide variety of features and options that may be employed. It is a formalization of weighting procedures such as nonresponse adjustment, poststratification, and weight trimming.

For nonresponse adjustments, GEM controls at the margins as opposed to controlling at the cell level, as weighting class adjustments. This approach allows more variables to be considered. GEM is designed so that the sum of the unadjusted weights for all eligible units equals the sum of the adjusted weights for respondents.

Extreme weights occur in the ELS:2002 data due to small probabilities of sample selection or due to weight adjustments. These extreme weights (either very small or very large) can significantly increase the variance of estimates. One way to account for this and decrease the variance is to trim and smooth extreme weights within prespecified domains. Note that trimming weights has the potential to increase bias. However, the increase in bias is often offset by the decrease in variance due to weight trimming. As a result, this reduces the mean square error of an estimate, defined as variance plus bias squared.

The innovation introduced in GEM is the ability to incorporate specific lower and upper bounds. An important application of this feature is to identify at each adjustment step an initial set of cases with extreme weights and to use specific bounds to exercise control over the final adjusted weights. Thus, there is built-in control for extreme weights in GEM.

GEM uses the median +/– X * IQR to identify extreme weights, where X is any number, typically between 2 and 3, and IQR is the interquartile range. There are also different points in the weight adjustment process during which weight trimming can occur. GEM has options to make adjustments for extreme weights as part of the nonresponse and as part of the poststratification. GEM adjusted for ELS:2002 second follow-up extreme weights during both nonresponse adjustment and during calibration. For GEM, a variable or set of variables is identified to be used to identify extreme weights within each level of the variable(s), and the variables race and school type were chosen. Prior to running GEM, the unweighted and weighted percentage of extreme weights was examined for four levels of race crossed with three levels of school type using various values to multiply by the IQR (2.0, 2.1, 2.2,…4.0), and multiples of the IQR were selected for each trimming process.

6.3.3.2 Predictor Variables for Nonresponse Models

In order to create weight adjustments that account for nonresponse, predictor variables must be incorporated into the modeling process. As the modeling process uses both respondents and nonrespondents, the information included in the nonresponse models must be known for both respondents and nonrespondents.

The second follow-up respondents include individuals who were base-year nonrespondents and include individuals who were first follow-up nonrespondents. Consequently, most information collected as part of the base-year and first follow-up surveys could not be used in the nonresponse adjustments. The variables used in the nonresponse models primarily consisted of sampling frame information, base-year sample school information, and some demographic characteristics. Table 61 lists all information that was used in at least one of the nonresponse models created for the second follow-up.

All school-level information was included in every nonresponse model and was only removed, where necessary, from those models in order to ensure model convergence. Because

the student-level information was not available for all second follow-up sample members, some information was used in some models but not in others. Details of the student-level information used in the various nonresponse models may be found in sections 6.3.5 and 6.3.6.

6.3.3.3 CHAID for Nonresponse Models

For those nonresponse adjustments that included interactions of the items listed in table 61, CHAID was performed on the predictor variables in order to detect important interactions for the logistic models used to produce nonresponse weight adjustment factors. The CHAID analysis divided the data into segments that differed with respect to the response variable (fielded, did not refuse, or respondent, depending on the model). The segmentation process first divided the sample into groups based on categories of the most significant predictor of response. It then split each of these groups into smaller subgroups based on other predictor variables. It also merged categories of a variable that were found to be insignificant. The splitting and merging process continued until no more statistically significant predictors were found or until some other stopping rule was met. The interactions from the final CHAID segments were then defined.

The interaction segments and all main effects were subjected to variable screening in the GEM logistic procedure. The initial model for a given adjustment step included all of the variables listed in table 61 that were available for respondents and nonrespondents and, where interaction terms were used, included the segments identified via CHAID. The most insignificant variables were deleted sequentially until the deletion of additional variables did not appreciably improve the UWE. Different bounds on the weight adjustments, depending on whether the weights were classified as extreme, were used to accomplish nonresponse adjustment, truncation, and smoothing in one step.

Table 61. Information used in nonresponse models: 2006

School-level information	Student-level information
School type	Student race/ethnicity
Metropolitan status	Student sex
Region	Student's native language
10th-grade enrollment	Family composition
Total enrollment	Parents' highest level of education
Number of minutes per class	Mother/female guardian's occupation
Number of class periods	Father/male guardian's occupation
Number of school days	Total family income from all sources
Percentage of students receiving free or reduced-price lunch	Socioeconomic status (SES)
Number of full-time teachers	G10COHRT—member of the sophomore cohort
Percentage of full-time teachers certified	G12COHRT—member of the senior cohort
Number of part-time teachers	Enrollment status
Number of different grades taught at the school	
School level	
Coeducational status	
Percentage of students with an Individualized Education Program	
Percentage of students with limited English proficiency	
Percentage of Hispanic 10th-grade students	
Percentage of Asian 10th-grade students	
Percentage of Black 10th-grade students	

NOTE: School-level information is from the base year (2002).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002" and "Second Follow-up, 2006."

6.3.4 Base Weight and Screening Adjustments

The base weight used to produce each of the second follow-up analysis weights was the first follow-up design weight, F1DWT. As described in the ELS:2002 base-year to first follow-up data file documentation, a school nonresponse adjustment (denoted WTADJ1) was applied to F1DWT for those first follow-up sample members who were part of the spring-term 12th-grade freshened sample in order to account for those schools that did not respond to the freshening process used in the first follow-up. The value of WTADJ1 was equal to 1 for those ELS:2002 sample members who were not part of the freshened sample of students added in the first follow-up. This same adjustment was applied to F1DWT and the resulting adjusted weight, denoted F2DWT, was taken as the second follow-up design weight. All second follow-up analysis weights were produced by applying a series of nonresponse and calibration adjustments to F2DWT.

As noted in section 6.3.3, some ELS:2002 sample members were not fielded as part of second follow-up data collection. These sample members included some first follow-up nonrespondents. Instead of assuming that these sample members would have retained their first follow-up status if they had been interviewed in the second follow-up, two ratio adjustments were created and applied to F2DWT in order to account for the likelihood that some of the first follow-up nonrespondents would have become out of scope for the second follow-up and to account for the likelihood that some out-of-scope cases would have become in scope for the second follow-up.

Since the number of ELS:2002 sample members not fielded for the second follow-up is small (less than 400) the resulting scope adjustments were very close to 1. The average ratio adjustment (denoted WTADJ2) for the first follow-up nonrespondents not fielded for the second follow-up was 1.0004. The average ratio adjustment (denoted WTADJ3) for the first follow-up out-of-scope cases not fielded for the second follow-up was 1.0006. The second follow-up adjusted, interim weight, F2IWT, calculated as:

F2IWT = F1DWT*WTADJ1*WTADJ2*WTADJ3

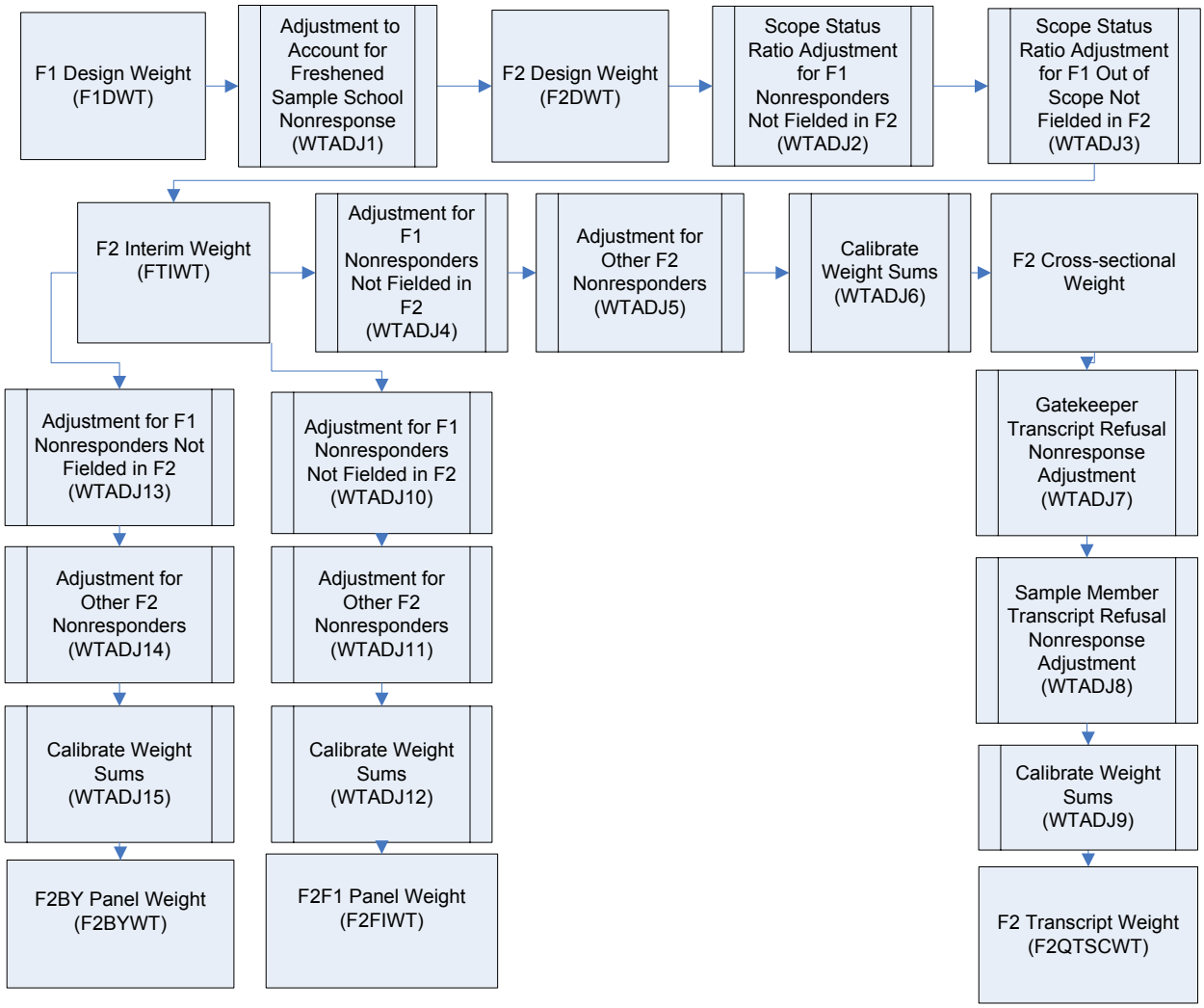
was used to produce each of the four second follow-up analysis weights. Subsequent adjustments to F2IWT varied by second follow-up analysis weight. The nonresponse and calibration adjustments applied to F2IWT to produce the second follow-up cross-sectional weights are described in section 6.3.5. The nonresponse and calibration adjustments applied to F2IWT to produce the second follow-up panel weights are described in section 6.3.6. Figure 13 summarizes the weight adjustments applied to the first follow-up design weight in order to produce the four second follow-up analysis weights.

6.3.5 Details of Weight Adjustments for Cross-sectional Weights

Two cross-sectional analysis weights were produced for the ELS:2002 second follow-up. The first cross-sectional weight was calculated for all sample members who fully or partially completed a second follow-up questionnaire. The second cross-sectional weight was calculated for all sample members who fully or partially completed a second follow-up questionnaire and for whom a transcript was collected as part of the first follow-up transcript study. The

nonresponse and calibration adjustments used to produce these two weights are described in sections 6.3.5.1 and 6.3.5.2.

Figure 13. Second follow-up weight adjustments: 2006



SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

6.3.5.1 Cross-sectional Weight: F2QWT

This second follow-up cross-sectional weight, F2QWT, was computed for those sample members who fully or partially completed the second follow-up questionnaire. Unlike prior rounds, prior-round questionnaire-incapable sample members who did not respond in the second follow-up were considered to be out of scope.

With a few exceptions, first follow-up eligible sample students remained eligible for the second follow-up sample. Students who died were out of scope for the second follow-up. Students who left the country, were unavailable for the duration of the study (e.g., in military boot camp), or were institutionalized were temporarily out of scope for the second follow-up, although they may be eligible in future rounds.

As noted in section 6.3.3, two nonresponse adjustments were created in order to account for nonresponse arising via two mechanisms:

- nonresponse arising from not fielding some first follow-up nonrespondents; and
- nonresponse resulting from fielded sample members not responding (either because they could not be contacted, could not be located, or refused to participate).

Also as noted in section 6.3.3, nonresponse resulting from the inability to locate/contact fielded sample members and nonresponse resulting from direct sample member refusal were treated as one nonresponse mechanism. The rationale for treating the reasons for nonresponse as one mechanism was based on the distribution of nonresponse cases. A review of the nonresponse cases indicated that the main reason for nonresponse was direct sample member refusal. A determination was made that the number of nonresponse cases associated with inability to locate or contact was not sufficient to warrant a separate nonresponse adjustment.

Weight adjustment for not-fielded cases. Some of the base-year nonrespondents were subsampled for inclusion in the first follow-up study; some of those base-year nonrespondents were nonrespondents or out of scope in the first follow-up. These sample members were not fielded in the second follow-up. Since the information available for the first follow-up nonrespondents not fielded for the second follow-up was limited, only a subset of the student-level information listed in table 61 was able to be used to create this nonresponse adjustment. In addition to all school-level variables listed in table 61, only student race/ethnicity and student sex were known for the first follow-up nonrespondents not fielded for the second follow-up.

A total of 23 variables were used as main effects in the GEM process. Additionally, as the number of first follow-up nonrespondents not fielded for the second follow-up was small, interactions of the main effects were not included in this first modeling process. The nonresponse adjustment factor resulting from this process is denoted WTADJ4.

The GEM process used to calculate nonresponse adjustments included a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed and smoothed weights. The values of the weight F2IWT were examined and extreme weights (3.8 percent unweighted and 13.6 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor, WTADJ4.

Table I-1 (appendix I) lists the final predictor variables used in the student nonresponse adjustment model that accounts for those first follow-up nonresponding sample members not fielded for the second follow-up. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.1 to 2.0 with a median of 1.1.

The temporary weight F2IWT*WTADJ4 was the input to the process used to calculate the nonresponse adjustment due to sample member refusal.

Weight adjustment for sample member nonresponse. Since the ELS:2002 sample members fielded for the second follow-up were base-year respondents, first follow-up respondents, or both, more student-level information could be used in the calculation of this

nonresponse adjustment than for the nonresponse adjustment WTADJ4. In addition to all school-level variables listed in table 61, all student-level variables except Enrollment Status were known for second follow-up respondents and second follow-up nonrespondents.

A total of 32 variables were used as main effects in the GEM process. These variables were also used in a CHAID analysis to determine important interactions for the nonresponse adjustment model. The nonresponse adjustment factor resulting from this process is denoted WTADJ5.

The GEM process used to calculate nonresponse adjustments included a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed and smoothed weights. The values of the weight F2IWT*WTADJ4 were examined and extreme weights (4.3 percent unweighted and 11.4 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor WTADJ5.

Table I-2 (appendix I) lists the final predictor variables (main effects and interactions) used in the student nonresponse adjustment model that accounts for those second follow-up fielded sample members who did not respond. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.7 to 2.7 with a median of 1.1.

The temporary weight F2IWT*WTADJ4*WTADJ5 was the input to the process used to calculate the calibration adjustment necessary to ensure that the second follow-up cross-sectional weight would preserve prior-round weight sums.

Weight adjustment used to calibrate weight sums. A weight adjustment factor was calculated using GEM to ensure that the second follow-up cross-sectional analysis weight preserved overall and marginal totals from prior rounds. The ELS:2002 sample members included in the weight calibration include second follow-up respondents and second follow-up out-of-scope sample members. In prior rounds, questionnaire-incapable members were considered respondents in the weight calibration, but in the second follow-up, questionnaire-incapable sample members were considered to be out of scope. Since these questionnaire-incapable members were included in the second follow-up calibration, the control totals used in the calibration process were derived from prior-round weight totals that include the questionnaire-incapable sample members. For the second follow-up cross-sectional weight F2QWT, control totals were calculated using the first follow-up expanded sample cross-sectional weight F1EXPWT.

Six key variables were used in the modeling process: Census region, School type, Sex, Race/ethnicity, 10th-grade cohort, and 12th-grade cohort. Interactions of 10th- and 12th-grade cohort with the other variables (Census region, School Type, Sex, and Race/ethnicity) were also included in the calibration model. The resulting calibration adjustment factor is denoted WTADJ6.

The GEM process used to calculate calibration adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2IWT*WTADJ4*WTADJ5 were examined and extreme

weights (3.2 percent unweighted and 9.4 percent weighted) were identified. The extreme weights were flagged and used to help produce the final calibration adjustment factor WTADJ6.

Table I-3 (appendix I) lists the final model variables (main effects and interactions) for which weight sums were preserved. This table also lists the control total and average weight adjustment by each level of each variable used in the calibration model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.1 to 1.5 with a median of 1.0.

The final second follow-up cross-sectional weight (F2QWT) is calculated as:

F2QWT = F1DWT*WTADJ1*WTADJ2*WTADJ3*WTADJ4*WTADJ5*WTADJ6.

Table 62 shows various statistical properties of the final second follow-up cross-sectional weight F2QWT.

Table 62. Statistical properties of cross-sectional weight F2QWT: 2006

Weight	F2QWT
Mean	240.7
Variance	26,560.5
Standard deviation	163
Coefficient of variation (x 100)	67.7
Minimum	5.4
Maximum	1,001.1
Skewness	1.0
Kurtosis	0.6
Sum	3,408,100
Number of cases	14,200

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

6.3.5.2 Cross-sectional Transcript Weight: F2QTSCWT

The second follow-up cross-sectional transcript weight (F2QTSCWT) was computed for those sample members who fully or partially completed a second follow-up questionnaire and for whom a first follow-up transcript was collected.

The second follow-up cross-sectional transcript weight was created by adjusting the second follow-up cross-sectional weight, F2QWT, described in section 6.3.5.1. Following the process developed in the first follow-up transcript study, three adjustments were applied to F2QWT. The first adjustment was a nonresponse adjustment used to account for those ELS:2002 sample members who did not have a transcript because a school or parent refused to provide the transcript. The second was a nonresponse adjustment used to account for those sample members who refused to allow their transcript to be collected. The third adjustment was used to calibrate weight sums of the cross-sectional transcript weight in order to preserve prior-round weight totals.

Weight adjustment for nonresponse due to gatekeepers. Since the cross-sectional transcript weight was created by applying adjustments to the second follow-up cross-sectional weight, all school- and student-level information listed in table 61 could be used in this nonresponse adjustment. The student-level Enrollment Status variable was incorporated into this

nonresponse adjustment as it was considered to be related to whether or not a transcript was available.

A total of 33 variables were used as main effects in the GEM process. These variables were also used in a CHAID analysis to determine important interactions for the nonresponse adjustment model. The nonresponse adjustment factor resulting from this process is denoted WTADJ7.

The GEM process used to calculate nonresponse adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2QWT were examined and extreme weights (1.7 percent unweighted and 4.3 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor WTADJ7.

Table I-4 (appendix I) lists the final predictor variables used in the student nonresponse adjustment model that accounts for the transcript nonresponse arising from gatekeeper refusal. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.7 to 2.9 with a median of 1.0.

The temporary weight F2QWT*WTADJ7 was the input to the process used to calculate the nonresponse adjustment necessary to account for transcript nonresponse arising from sample member refusal.

Weight adjustment due to sample member refusal. The variables used to calculate the nonresponse adjustment to account for gatekeeper refusal were also used in the process to calculate a nonresponse adjustment to account for those sample members who refused permission to include transcript data with the first follow-up data. All variables listed in table 61 were used to calculate the nonresponse adjustment factor for sample member refusal.

A total of 33 variables were used as main effects in the GEM process. These variables were also used in a CHAID analysis to determine important interactions for the nonresponse adjustment model. The nonresponse adjustment factor resulting from this process is denoted WTADJ8.

The GEM process used to calculate nonresponse adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the temporary weight F2QWT*WTADJ7 were examined and extreme weights (1.0 percent unweighted and 2.9 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor WTADJ8.

Table I-5 (appendix I) lists the final predictor variables used in the student nonresponse adjustment model that accounts for the transcript nonresponse arising from sample member refusal. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.4 to 2.9 with a median of 1.0.

The temporary weight F2QWT*WTADJ7*WTADJ8 was the input to the process used to calculate the calibration adjustment necessary to ensure that the second follow-up cross-sectional transcript weight would preserve prior-round weight sums.

Weight adjustment used to calibrate weight sums. A weight adjustment factor was calculated using GEM to ensure that the second follow-up cross-sectional transcript weight preserved overall and marginal totals from prior rounds. The ELS:2002 sample members included in the weight calibration include second follow-up respondents and out-of-scope sample members. In prior rounds, questionnaire-incapable members were considered respondents in the weight calibration but, in the second follow-up, questionnaire-incapable sample members were considered to be out of scope. Since these questionnaire-incapable members were included in the second follow-up calibration, the control totals used in the calibration process were derived from prior-round weight totals that include the questionnaire-incapable sample members. For the second follow-up cross-sectional transcript weight F2QTSCWT, control totals were calculated using the first follow-up expanded sample cross-sectional weight F1EXPWT.

Six key variables were used in the modeling process: Census region, School type, Sex, Race/ethnicity, 10th-grade cohort, and 12th-grade cohort. Interactions of 10th- and 12th-grade cohort with the other variables (Census region, School Type, Sex, and Race/ethnicity) were also included in the calibration model. The resulting calibration adjustment factor is denoted WTADJ9.

The GEM process used to calculate calibration adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2QWT*WTADJ7*WTADJ8 were examined and extreme weights (1.2 percent unweighted and 3.7 percent weighted) were identified. The extreme weights were flagged and used to help produce the final calibration adjustment factor WTADJ9.

Table I-6 (appendix I) lists the final model variables, main effects, and interactions for which weight sums were preserved. This table also lists the control total and average weight adjustment by each level of each variable used in the calibration model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.4 to 1.4 with a median of 1.0.

The final second follow-up cross-sectional transcript weight (F2QTSCWT) is calculated as:

F2QTSCWT = F2QWT*WTADJ7*WTADJ8*WTADJ9.

Table 63 shows various statistical properties of the final second follow-up cross-sectional transcript weight F2QTSCWT.

Table 63. Statistical properties of the cross-sectional transcript weight F2QTSCWT: 2006

Weight	F2QTSCWT
Mean	262.0
Variance	33,044.6
Standard deviation	181.8
Coefficient of variation (x 100)	69.4
Minimum	5.4
Maximum	1,031.1
Skewness	1.0
Kurtosis	0.7
Sum	3,408,100
Number of cases	13,000

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

6.3.6 Details of Weight Adjustments for Panel Weights

Two panel analysis weights were produced for the ELS:2002 second follow-up. The first panel weight, covering the first and second follow-up rounds, was calculated for all sample members who:

- fully or partially completed a second follow-up questionnaire and fully or partially completed a first follow-up questionnaire; or
- fully or partially completed a second follow-up questionnaire and were questionnaire-incapable in the first follow-up.

The second panel weight, covering the base-year and second follow-up rounds, was calculated for all sample members who:

- fully or partially completed a second follow-up questionnaire and fully or partially completed a base-year questionnaire; or
- fully or partially completed a second follow-up questionnaire and were questionnaire incapable in the base year; or
- were base-year nonrespondents who responded in the first and second follow-ups and for whom base-year classification information was collected in the first follow-up, when their test scores were also imputed.

The nonresponse and calibration adjustments used to produce these two weights are described in sections 6.3.6.1 and 6.3.6.2, respectively.

6.3.6.1 First Follow-up to Second Follow-up Panel Weight: F2F1WT

This second follow-up panel weight was computed for those sample members who fully or partially completed the second follow-up questionnaire and responded[79] in the first follow-up. Unlike prior rounds, questionnaire-incapable sample members who did not respond in the second follow-up were considered to be out of scope.

[79] Such sample members included first follow-up respondents and first follow-up questionnaire-incapable sample members.

The sample members who were assigned a first follow-up to second follow-up analysis weight were a subset of those sample members who had a second follow-up cross-sectional weight. Second follow-up respondents who were nonrespondents in the first follow-up or were out of scope in the first follow-up were not considered panel respondents.

The nonresponse adjustments used to create F2F1WT accounted for the same two nonresponse mechanisms used in the adjustment process to create F2QWT and are described in section 6.3.5.1. In particular, two nonresponse adjustments were created in order to account for nonresponse arising via two mechanisms:

- nonresponse arising from not fielding some first follow-up nonrespondents; and
- nonresponse resulting from fielded sample members not responding (because they could not be contacted, could not be located, or refused to participate).

Also as noted in section 6.3.3, nonresponse resulting from the inability to locate/contact fielded sample members and nonresponse resulting from direct sample member refusal were treated as one nonresponse mechanism. The rationale for treating these reasons for nonresponse as one mechanism was based on the distribution of nonresponse cases. A review of the nonresponse cases indicated that the main reason for nonresponse was direct sample member refusal. A determination was made that the number of nonresponse cases associated with inability to locate was not sufficient to warrant a separate nonresponse adjustment.

Weight adjustment for not fielding cases. Some of the base-year nonrespondents were subsampled for inclusion in the first follow-up study and some were nonrespondents or out of scope in the first follow-up. These sample members were not fielded in the second follow-up. Since the information available for the first follow-up nonrespondents not fielded for the second follow-up was limited, only a subset of the student-level information listed in table 61 could be used to create this nonresponse adjustment. In addition to all school-level variables listed in table 61, only student race/ethnicity and student sex were known for the first follow-up nonrespondents not fielded for the second follow-up.

A total of 23 variables were used as main effects in the GEM process. Additionally, as the number of first follow-up nonrespondents not fielded for the second follow-up was small, interactions of the main effects were not included in this first modeling process. The nonresponse adjustment factor resulting from this process is denoted WTADJ10.

The GEM process used to calculate nonresponse adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2IWT were examined and extreme weights (3.8 percent unweighted and 13.6 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor WTADJ10.

Table I-7 (appendix I) lists the final predictor variables used in the student nonresponse adjustment model that accounts for those first follow-up nonresponding sample members not fielded for the second follow-up. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.1 to 2.0 with a median of 1.1.

The temporary weight F2IWT*WTADJ10 was the input to the process used to calculate the nonresponse adjustment accounting for sample member refusal.

Weight adjustment for sample member nonresponse. Since the ELS:2002 sample members fielded for the second follow-up were base-year respondents, first follow-up respondents, or both, more student-level information could be used in the calculation of this nonresponse adjustment than for the nonresponse adjustment WTADJ10. In addition to all school-level variables listed in table 61, all student-level variables except Enrollment Status were known for second follow-up respondents and second follow-up nonrespondents.

A total of 32 variables were used as main effects in the GEM process. These variables were also used in a CHAID analysis to determine important interactions for the nonresponse adjustment model. The nonresponse adjustment factor resulting from this process is denoted WTADJ11.

The GEM process used to calculate nonresponse adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2IWT*WTADJ10 were examined and extreme weights (4.3 percent unweighted and 11.6 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor WTADJ11.

Table I-8 (appendix I) lists the final predictor variables, main effects, and interactions used in the student nonresponse adjustment model that accounts for those second follow-up fielded sample members who did not respond. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.7 to 2.4 with a median of 1.2.

The temporary weight F2IWT*WTADJ10*WTADJ11 was the input to the process used to calculate the calibration adjustment necessary to ensure that prior-round weight sums were preserved.

Weight adjustment used to calibrate weight sums. A weight adjustment factor was calculated using GEM to ensure that the second follow-up panel weight F2F1WT preserved overall and marginal totals from prior rounds. The ELS:2002 sample members included in the weight calibration include second follow-up respondents and second follow-up out-of-scope sample members. In prior rounds, questionnaire-incapable members were considered respondents in the weight calibration but, in the second follow-up, questionnaire-incapable sample members were considered to be out of scope. Since these questionnaire-incapable members were included in the second follow-up calibration, the control totals used in the calibration process were derived from prior-round weight totals that include the questionnaire-incapable sample members. For the second follow-up panel weight F2F1WT, control totals were calculated using the first follow-up expanded sample cross-sectional weight F1EXPWT.

Six key variables were used in the modeling process: Census region, School type, Sex, Race/ethnicity, 10th-grade cohort, and 12th-grade cohort. Interactions of 10th- and 12th-grade cohort with the other variables (Census region, School type, Sex, and Race/ethnicity) were also included in the calibration model. The resulting calibration adjustment factor is denoted WTADJ12.

The GEM process used to calculate calibration adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2IWT*WTADJ10*WTADJ11 were examined and extreme weights (2.9 percent unweighted and 8.5 percent weighted) were identified. The extreme weights were flagged and used to help produce the final calibration adjustment factor WTADJ12.

Table I-9 (appendix I) lists the final model variables, main effects, and interactions for which weight sums were preserved. This table also lists the control total and average weight adjustment by each level of each variable used in the calibration model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.1 to 1.5 with a median of 1.00.

The final panel weight F2F1WT is calculated as:

F2F1WT=F1DWT*WTADJ1*WTADJ2*WTADJ3*WTADJ10*WTADJ11*WTADJ12.

Table 64 shows various statistical properties of the final second follow-up panel weight F2F1WT.

Table 64. Statistical properties of panel weight F2F1WT: 2006

Weight	F2F1WT
Mean	254.0
Variance	30,503.1
Standard deviation	174.7
Coefficient of variation (x 100)	68.8
Minimum	5.6
Maximum	1,041.3
Skewness	1.0
Kurtosis	0.6
Sum	3,394,800
Number of cases	13,400

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

6.3.6.2 Base-Year to Second Follow-up Panel Weight: F2BYWT

The second follow-up panel weight (F2BYWT) was computed for those sample members who fully or partially completed the second follow-up questionnaire and responded[80] in the base-year round. Unlike prior rounds, questionnaire-incapable sample members who did not respond in the second follow-up were considered to be out of scope.

The sample members who were assigned a base-year to second follow-up analysis weight were a subset of those sample members who had a second follow-up cross-sectional weight. The nonresponse adjustments used to create F2BYWT accounted for the same two nonresponse mechanisms used in the adjustment process used to create F2QWT and F2F1WT and are described in sections 6.3.5.1 and 6.3.6.1. In particular, two nonresponse adjustments were created in order to account for nonresponse arising via two mechanisms:

- nonresponse arising from not fielding some first follow-up nonrespondents; and

[80] Such sample members included base-year respondents and base-year questionnaire-incapable sample members.

- nonresponse resulting from fielded sample members not responding (either because they could not be contacted, could not be located, or refused to participate).

Also as noted in section 6.3.3, nonresponse resulting from the inability to locate/contact fielded sample members and nonresponse resulting from direct sample member refusal were treated as one nonresponse mechanism. The rationale for treating these reasons for nonresponse as one mechanism was based on the distribution of nonresponse cases. A review of the nonresponse cases indicated that the main reason for nonresponse was direct sample member refusal. A determination was made that the number of nonresponse cases associated with inability to locate was not sufficient to warrant a separate nonresponse adjustment.

Weight adjustment for no-field cases. Some of the base-year nonrespondents were subsampled for inclusion in the first follow-up study and some were nonrespondents or out of scope in the first follow-up. These sample members were not fielded in the second follow-up. Since the information available for these cases was limited, only a subset of the student-level information listed in table 65 could be used to create this nonresponse adjustment. In addition to all school-level variables listed in table 65, only student race/ethnicity and student sex were known for the base-year nonrespondents who were also first follow-up nonrespondents not fielded for the second follow-up.

A total of 23 variables were used as main effects in the GEM process. Additionally, as the number of first follow-up nonrespondents not fielded for the second follow-up was small, interactions of the main effects were not included in this first modeling process. The nonresponse adjustment factor resulting from this process is denoted WTADJ13.

The GEM process used to calculate nonresponse adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2IWT were examined and extreme weights (3.7 percent unweighted and 13.6 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor WTADJ13.

Table I-10 (appendix I) lists the final predictor variables used in the student nonresponse adjustment model that accounts for those base-year nonrespondents who were first follow-up nonrespondents not fielded for the second follow-up. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.1 to 1.9 with a median of 1.1.

The temporary weight F2IWT*WTADJ13 was the input to the process used to calculate the nonresponse adjustment necessary to account for nonresponse among fielded cases in the second follow-up.

Weight adjustment sample member nonresponse. Since the ELS:2002 sample members fielded for the second follow-up were base-year respondents, first follow-up respondents, or both, more student-level information could be used in the calculation of this nonresponse adjustment than for the nonresponse adjustment WTADJ13. In addition to all school-level variables listed in table 65, all student-level variables except Enrollment Status were known for second follow-up respondents and second follow-up nonrespondents. The variable G10COHRT

was not used in this weight adjustment process since all respondents and nonrespondents for this panel weight are in the 10th-grade cohort.

A total of 31 variables were used as main effects in the GEM process. These variables were also used in a CHAID analysis to determine important interactions for the nonresponse adjustment model. The nonresponse adjustment factor resulting from this process is denoted WTADJ14.

The GEM process used to calculate nonresponse adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2IWT*WTADJ13 were examined and extreme weights (4.14 percent unweighted and 11.1 percent weighted) were identified. The extreme weights were flagged and used to help produce the final nonresponse adjustment factor WTADJ14.

Table I-11 (appendix I) lists the final predictor variables, main effects, and interactions used in the student nonresponse adjustment model that accounts for those second follow-up fielded sample members who did not respond. This table also lists the number of respondents, the weighted response rate, and the average weight adjustment by each level of each predictor variable included in the final nonresponse model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.1 to 1.8 with a median of 1.1.

The temporary weight F2IWT*WTADJ13*WTADJ14 was the input to the process used to calculate the calibration adjustment necessary to ensure that prior-round weight sums were preserved.

Weight adjustment used to calibrate weight sums. A weight adjustment factor was calculated using GEM to ensure that the second follow-up panel weight F2BYWT preserved overall and marginal totals from prior rounds. The ELS:2002 sample members included in the weight calibration include second follow-up respondents and second follow-up out-of-scope sample members. In prior rounds, questionnaire-incapable members were considered respondents in the weight calibration but, in the second follow-up, questionnaire-incapable sample members were considered to be out of scope. Since these questionnaire-incapable members were included in the second follow-up calibration, the control totals used in the calibration process were derived from prior-round weight totals that include the questionnaire-incapable sample members. For the second follow-up panel weight F2BYWT, control totals were calculated using the first follow-up expanded sample cross-sectional weight BYEXPWT.

Four key variables were used in the modeling process: Census region, School type, Sex, and Race/ethnicity. Since all sample members who received a F2BYWT are in the 10th-grade cohort, there was no need to include the variable G10COHRT in the calibration model. Additionally, since the set of ELS:2002 sample members eligible for this panel weight is not representative of the 12th-grade cohort, control totals for the 12th-grade cohort were not preserved. The resulting calibration adjustment factor is denoted WTADJ15.

The GEM process used to calculate calibration adjustments includes a trimming process and the final weight adjustment factor is calculated in such a fashion as to produce trimmed weights. The values of the weight F2IWT*WTADJ13*WTADJ14 were examined and extreme weights (3.2 percent unweighted and 9.2 percent weighted) were identified. The extreme weights were flagged and used to help produce the final calibration adjustment factor WTADJ15.

Table I-12 (appendix I) lists the final model variables for which weight sums were preserved. This table also lists the control total and average weight adjustment by each level of each variable used in the calibration model. While the average adjustment factor, by variable level, was generally near 1, the individual student-level adjustment factors varied from 0.1 to 1.1 with a median of 1.0.

The final panel weight F2BYWT is calculated as:

F2BYWT=F1DWT*WTADJ1*WTADJ2*WTADJ3*WTADJ13*WTADJ14*WTADJ15.

Table 65 shows various statistical properties of the final second follow-up panel weight F2BYWT.

Table 65. Statistical properties of panel weight F2BYWT: 2006

Weight	F2BYWT
Mean	239.4
Variance	26,188.0
Standard deviation	161.8
Coefficient of variation (x 100)	67.6
Minimum	5.3
Maximum	793.0
Skewness	0.9
Kurtosis	0.5
Sum	3,357,400
Number of cases	14,000

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

6.3.7 BRR Weights for the Data Analysis System

Four sets of 200 BRR replicate weights were computed as an alternative variance estimation procedure because NCES's DAS requires BRR weights for variance estimation. The 200 replicates were constructed so that there are a sufficient number of replicates for the new regression feature of the DAS. The four sets correspond to the four weights described in section 6.3.2:

- F2Q1—F2Q200;
- F2TRS1—F2TRS200;
- F2BYP1—F2BYP200; and
- F2F1P1—F2F1P200.

The second follow-up replicate weights were computed in a similar manner to those computed for the base year and first follow-up.

The BRR procedure is an alternative variance estimation procedure that computes the variance based on a balanced set of pseudoreplicates. The BRR variance estimation process involves modeling the design as if it were a two-PSU-per-stratum design. Variances were calculated using a random group type of variance estimation procedure, with a balanced set of 200 replicates as the groups. Balancing was done by using an orthogonal matrix (200 x 200

Hadamard matrix) and allows the use of less than the full set of 2^L possible replicates, where L is the number of analysis strata. To achieve full orthogonal balance, the number of BRR strata needs to be less than the number of replicates. Therefore, we created 200 replicates in 199 strata. Section 6.3.7.1 describes the strata and PSUs (replicates) that were created for the base-year replicate weights and used again in the first and second follow-ups. Section 6.3.7.2 describes the weight adjustments made for the second follow-up, and section 6.3.7.3 summarizes the results of the replicate weighting.

6.3.7.1 Strata and PSUs

For Taylor series variance estimation, 361 analysis strata containing responding schools were created from the 96 sampling strata based on the sample design. In order to replicate the school weight, it is necessary for the BRR strata to contain all sample schools (respondents and nonrespondents). For the base year, 594 analysis strata were formed for the purpose of computing school-level Taylor series variance estimates. We collapsed these 594 analysis strata into 199 BRR strata. We estimated the base-year expected sample size for each sample school in the 594 strata and then collapsed strata randomly across size groups (small, medium, large) so that the 199 strata have approximately equal sizes. Collapsing randomly allows schools of different types, regions, or urbanicities to be together in a stratum. This provides more degrees of freedom for variance estimation for domains and helps obtain more accurate variance estimates within domains. Within the 199 BRR strata, there are two PSUs. Each school in a stratum was randomly assigned to one of the two PSUs.

The strata were randomly assigned to the rows of the Hadamard matrix. The 200 columns of the matrix are the replicates. Within each stratum, the matrix contains values of +1 and -1; one PSU was randomly assigned +1 and the other PSU was assigned -1. For PSUs with a value of +1, the school base (sampling) weight was multiplied by 2 to create the initial BRR weight, otherwise the school base weight was multiplied by zero. Approximately half of the schools in each of the 200 replicates have initial BRR weights of zero and the other half have initial BRR weights double the initial base weight.

6.3.7.2 Weight Adjustments

While both Taylor series and BRR variance estimation methods reflect the increase in variance due to unequal weighting, the BRR weights can also be designed to reflect the variance impact (increase or decrease) of the weight adjustment process. The impact of the weight adjustment process is captured by repeating nonresponse adjustment and calibration processes on each BRR half sample.

The F2 replication process mirrored the F2 analysis weight construction, and the design weight was the F1 replicate design weight. All F2 weight adjustments were replicated, including the adjustment for unknown eligibility, two nonresponse adjustments, and calibration. The original F2 nonresponse and calibration models were used initially for each of the 200 replicates. However, some of the models did not converge for some replicates, so variables were deleted one by one from the models until convergence was achieved. The variables deleted were those that seemed to be causing the convergence problems, as long as they were not key design variables. The weight distribution was calibrated to the F1 weight sums. Since the F2 weights were not poststratified to external (known) totals, the estimates could legitimately reflect some variation in base-year totals due to sampling variability. To recognize the calibration to F1, each

half sample was calibrated to F1 half sample replicate weight sums rather than calibrated to F1 full sample analysis weight sums.

6.3.7.3 Results

When weights are adjusted by poststratification to align sample estimates with certain "known" population totals called controls, the sampling variance for estimates of the controls goes to zero, and the variance for related statistics is expected to be reduced. Repeating the poststratification (to the common "known" set of external totals) step on each half sample replicate ensures that the variance estimates for the control total estimates are zero and is expected to reduce the variance estimates for statistics correlated with the totals. However, when the calibration is to previous round half sample data, such as in the F2, the variance estimates for the control total estimates are not zero. This is because the control total for each replicate is different, hence there is variance between replicates.

Using the set of variables used to compute the design effects (see section 6.4), standard errors were computed using both the Taylor series and BRR variance estimation methods. Taylor series variance estimates were computed using the four F2 analysis weights, and the BRR variance estimation used the four sets of F2 BRR weights. For each of the four comparisons between the two methods, the Taylor series standard error was less than the BRR standard error for about 80 percent of the variables analyzed. Since BRR takes into account the variance due to weight adjustments, these results are expected.

6.3.8 Quality Control

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

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

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

6.4 Second Follow-up Standard Errors and Design Effects

6.4.1 Standard Errors

For probability-based sample surveys, most estimates are nonlinear statistics. For example, a mean or proportion, which is expressed as $\Sigma wy/\Sigma w$,[81] is nonlinear because the denominator is a survey estimate of the (unknown) population total. In this situation, the

[81] Where w is the sample weight, and y is a 0/1 variable indicating whether a certain characteristic is present for the sample member.

variances of the estimates cannot be expressed in closed form. One common procedure for estimating variances of survey statistics is the Taylor series linearization procedure. This procedure takes the first-order Taylor series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. Woodruff presented the mathematical formulation of this procedure (Woodruff 1971). The variance estimation must also take into account stratification and clustering. There are other variance estimation procedures, such as jackknife and BRR. Taylor series estimation was used for the base year and first follow-up and also used for the second follow-up. BRR weights were produced for the second follow-up for use in the ELS:2002/06 DAS.

Variance estimation procedures assumed a with-replacement design at the first stage of sampling. Because school sampling rates were moderately low, this assumption yields estimates that are only slightly biased in the positive direction. For stratified multistage surveys and a with-replacement sample design, the Taylor series procedure requires the specification of analysis strata and analysis PSUs. The base-year sampling design employed 96 sampling strata and 752 primary sampling units. Given that the school sample was selected using probability with minimum replacement, for variance estimation in the base year, variance estimation strata were formed consisting of two PSUs per stratum (Chromy 1981). Some 361 analysis strata, containing two PSUs per stratum, were formed by grouping together the 752 sampling PSUs. The responding schools were sorted within sampling strata in the same order as was used for sampling, and then adjacent analysis PSUs were paired to form analysis strata. However, whenever there was an odd number of schools in a sampling stratum, an analysis stratum with three PSUs would be formed. The same analysis strata and PSUs as in the base year were used in the first follow-up and in the second follow-up.

As described in chapter 3, the ELS:2002 base-year sampling design was a stratified two-stage design. A stratified sample of schools was selected with probabilities proportional to a composite measure of size at the first stage, and a stratified systematic sample of students was selected from sample schools at the second stage. At the first stage, the school sampling rates varied considerably by school sampling strata. At the second stage, Asian and Hispanic students were sampled at higher rates than other students. Because of this complex sampling design, statistical analyses should be conducted using software that properly accounts for the complex survey design.

Many commonly used statistical computing packages assume that the data were obtained from a simple random sample; that is, they assume that the observations are independent and identically distributed. When the data have been collected using a complex sampling design, the simple random sampling assumption usually leads to an underestimate of the sampling variance, which would lead to artificially small confidence intervals and liberal hypothesis test results (i.e., rejecting the null hypothesis when it is in fact true more often than indicated by the nominal Type I error level) (Carlson, Johnson, and Cohen 1993).

Statistical strategies that have been developed to address this issue include first-order Taylor series expansion of the variance equation, balanced repeated replication, and the jackknife approach (Wolter 2007). Special-purpose software packages that have been developed for analysis of complex sample survey data include SUDAAN, WesVar, and Stata. Evaluations of the relative performances of these packages are reported by Cohen (1997).

- SUDAAN is a commercial product developed by RTI International; information regarding the features of this package and its lease terms is available from the website http://www.rti.org/sudaan.
- WesVar is a product of Westat, Inc.; information regarding the features of this package and its lease terms is available from the website http://www.westat.com/wesvar.
- Information regarding the features of Stata and its lease terms is available from the website http://www.stata.com.
- In addition to the variance estimation packages noted above, the American Institutes for Research has developed the AM Statistical Software. AM software can be downloaded for free from the following website: http://am.air.org/.

Following is an example of generic SUDAAN code to produce estimates and standard errors using Taylor series, followed by an example from Stata. The symbols /* and */ in the code indicate the beginning and end of a comment. Note that the dataset must be sorted by analysis strata and analysis PSUs before analyzing the data in SUDAAN.

```
proc descript data=/* insert filename*/ design=wr;
nest analstr analpsu; /* these variables are the analysis strata and analysis PSUs,
respectively */
weight F2QWT;
var /*insert variables*/;
subpopn /* insert domain of interest if domain is a subset of students*/;
print nsum mean semean / style=nchs;
run;
```

Stata code is as follows:

```
drop _all

set memory 18000

use "/* insert filename */", clear


sort analstr analpsu /* these variables are the analysis strata and analysis PSUs,
respectively */


 svyset analpsu [pweight=f2qwt], strata(analstr)

svy: tab /*insert variables*/, subpop (name of domain) row se
```

The above reflects the version 9 command structure; earlier versions of Stata require the following syntax:

```
svyset [pweight=f2qwt], strata(analstr) psu(analpsu)

svytab /*insert variables*/, subpop (name of domain) row se
```

6.4.2 Design Effects

The impact of the departures of the ELS:2002 complex sample design from a simple random sample design on the precision of sample estimates can be measured by the design effect. The design effect is the ratio of the actual variance of the statistic to the variance that would have been obtained had the sample been a simple random sample. The design standard errors will be different from the standard errors that are based on the assumption that the data are from a simple random sample. The ELS:2002 sample departs from the assumption of simple random sampling in three major respects: student samples were stratified by student characteristics, students were selected with unequal probabilities of selection, and the sample of students was clustered by school. A simple random sample is, by contrast, unclustered and not stratified. Additionally, in a simple random sample, all members of the population have the same probability of selection. Generally, clustering and unequal probabilities of selection increase the variance of sample estimates relative to a simple random sample, and stratification decreases the variance of estimates.

Standard errors and design effects were computed for all respondents. Standard errors and design effects were computed for 30 means and proportions overall for all respondents and for subgroups of all respondents. The subgroups are similar to those used in NELS:88, the ELS:2002 base year, and the ELS:2002 first follow-up:

- sex (male and female);
- race/ethnicity (Asian/Pacific Islander, Black, Hispanic, White/other, multiracial);
- school type (public, Catholic, and other private);
- SES (lowest quarter, middle two quarters, and highest quarter); and
- postsecondary enrollment (ever enrolled in a postsecondary institution, never enrolled in a postsecondary institution).

It is important to compare design effects across cohorts (e.g., ELS:2002 versus NELS:88), so table 5.3.1 from the Methodology Report: NELS:88 Third Follow-Up (Haggerty et al. 1996) was initially used to help guide the items picked. However, the ELS:2002 items chosen differ quite a bit from the items used in constructing design effects for NELS:88 as there were substantial differences in the types and composition of variables produced in each study. Nonetheless, the items chosen are a good representation of the different items in the ELS:2002 second follow-up survey questionnaire. These items should provide a range of design effects that will give a reasonable average for both the entire sample and for analytically important subgroups. However, because item matching with NELS:88 was difficult, the ELS:2002 design

effects may not be comparable with the NELS:88 repeated design effects. Ideally, one would like to compare exact items between survey systems. Table 66 lists the 30 items chosen for computing design effects for all respondents and subgroups. For categorical variables, the item value corresponding to the category of interest is listed.

Table 66. Items chosen for computing design effects for all respondents and subgroups: 2006

Survey item	Variable name	Item value[1]
Ever dropped out	F2EVERDO	1
Fall 2003–Summer 2004 high school graduate	F2HSSTAT	1
Received GED or other equivalency	F2HSSTAT	6
Ever applied to a postsecondary school	F2EVRAPP	1
Meet with advisor about academic plans often	F2B18B	3
Participate in other extracurricular activities often	F2B18G	3
Postsecondary education paid with grants/scholarships	F2B25A	1
Expect to finish college, but not advanced degree	F2STEXP	6
Ever held a job since leaving high school	F2EVRJOB	1
First job is working for an employer	F2C07	1
Current employer offers health insurance	F2C21	1
At age 30 expects to have a job as a laborer	F2OCC30	5
At age 30 expects to have a job as a manager	F2OCC30	6
At age 30 expects to have a job in the military	F2OCC30	7
At age 30 expects to have a professional job (group *a*)	F2OCC30	9
At age 30 expects to have a sales job	F2OCC30	13
At age 30 expects to have a job as a school teacher	F2OCC30	14
College degree but not advanced degree needed for job at age 30	F2C41	6
Respondent's current marital status is single	F2D01	1
Respondent's current marital status is married	F2D01	2
Number of friends or roommates living with respondent	F2D08C	Continuous
Number of siblings living with respondent	F2D08D	Continuous
Respondent lives in school-provided housing in spring 2006	F2D07	1
Respondent performed community service in past 2 years	F2D09	1
Volunteered with school/community organizations	F2D10B	1
Volunteered with church-related group	F2D10D	1
Voted in 2004 presidential election	F2D13	1
Respondent served in military	F2D14	1
Respondent's parent/guardian divorced in last 2 years	F2D15A	1
Respondent's parent/guardian lost job in last 2 years	F2D15B	1

[1] For categorical variables, the item value corresponds to the category of interest, and for continuous variables, the item value is indicated as continuous.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

The variables used were the versions after imputation (see section 6.5), and all variables used were after disclosure avoidance (see section 6.6). For all respondents, the standard errors and design effects were calculated using both the cross-sectional weight (F2QWT) and the panel weight (F2F1WT). When using the panel weight, only panel respondents were included. The difference between the cross-sectional and panel respondents is that first follow-up nonrespondents who were second follow-up respondents are cross-sectional respondents but are not panel respondents.

Appendix J contains tables of design effects for all respondents. Each table includes the survey item (or composite variable), variable name and value, percent estimate, design standard error, simple random sample standard error, sample size (N), design effect (DEFF), and square root of the design effect (DEFT). Tables 67 and 68 summarize the average DEFFs and DEFTs for the full sample and panel sample, respectively, for all respondents and each subgroup. The reader should note that the mean DEFTs reported in tables 67 and 68 were not calculated directly from the mean DEFF but, rather, are based on the summary statistics from the tables in appendix J.

Table 67. Mean design effects and root design effects for the second follow-up full sample, by selected characteristics: 2006

Characteristic	Mean design effect	Mean root design effect
All respondents	1.90	1.37
Male	1.65	1.28
Female	1.71	1.30
American Indian or Alaska Native	1.39	1.17
Asian or Pacific Islander	1.53	1.23
Black or African American	1.44	1.20
Hispanic or Latino	1.48	1.21
White and all other races[1]	1.74	1.31
More than one race	1.62	1.27
Public schools	1.67	1.28
Catholic schools	1.63	1.26
Other private Schools	2.39	1.50
Low socioeconomic status (SES)	1.46	1.21
Middle SES	1.58	1.25
High SES	1.76	1.32
Ever enrolled in postsecondary	1.78	1.33
Never enrolled in postsecondary	1.44	1.20

[1] "White and all other races" is predominantly White, with a very small number of individuals from other race categories.

NOTE: The mean root design effect was not calculated directly from the mean design effect but, rather, is the average root design effect over selected items. See appendix J of this document for more information.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Table 68. Mean design effects and root design effects for the second follow-up panel sample, by selected characteristics: 2006

Characteristic	Mean design effect	Mean root design effect
All respondents	1.90	1.37
Male	1.66	1.29
Female	1.74	1.31
American Indian or Alaska Native	1.47	1.20
Asian or Pacific Islander	1.53	1.23
Black or African American	1.44	1.20
Hispanic or Latino	1.45	1.20
White and all other races[1]	1.75	1.32
More than one race	1.67	1.29
Public schools	1.66	1.28
Catholic schools	1.60	1.25
Other private schools	2.30	1.47
Low socioeconomic status (SES)	1.44	1.20
Middle SES	1.61	1.26
High SES	1.77	1.33
Ever enrolled in postsecondary	1.83	1.35
Never enrolled in postsecondary	1.43	1.20

[1] "White and all other races" is predominantly White, with a very small number of individuals from other race categories.

NOTE: The mean root design effect was not calculated directly from the mean design effect but, rather, is the average root design effect over selected items. See appendix J of this document for more information.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Table 69 shows the design effects from the base-year and first follow-up for subgroups. The second follow-up design effects are lower for all respondents and for all of the common subgroups used in design effects calculations than the base-year and first follow-up design effects.

The smaller design effects in the second follow-up compared with those in the base year and first follow-up may be due to the general tendency in longitudinal studies for design effects to lessen over time, as dispersion reduces the original clustering. In the second follow-up, almost all sample members had left the base-year school, the clusters of students within schools dispersed to an extent. Social characteristics of the sample members potentially varied to a greater extent as the clusters dispersed.

Table 69. Mean design effects for base-year and first follow-up student questionnaire data, by selected characteristics: 2002 and 2004

Group	Mean design effect base year	Mean design effect first follow-up full sample	Mean design effect first follow-up panel sample
All students	2.35	2.26	2.23
Dropouts	†	1.31	1.31
Male	1.90	1.90	1.88
Female	2.01	1.94	1.93
American Indian or Alaska Native	1.42	1.51	1.50
Asian or Pacific Islander	2.27	2.14	2.17
Black or African American	1.67	1.49	1.49
Hispanic or Latino	1.82	1.59	1.60
More than one race	1.63	1.71	1.70
White and all other races[1]	2.03	1.84	1.83
Public schools	2.07	1.97	1.94
Catholic schools	2.43	2.25	2.25
Other private schools	3.53	3.02	3.00
Low socioeconomic status (SES)	1.70	1.66	1.64
Middle SES	1.73	1.68	1.67
High SES	1.99	1.91	1.92
Urban	2.88	2.85	2.80
Suburban	2.15	2.08	2.08
Rural	1.94	1.71	1.71

† Not applicable.

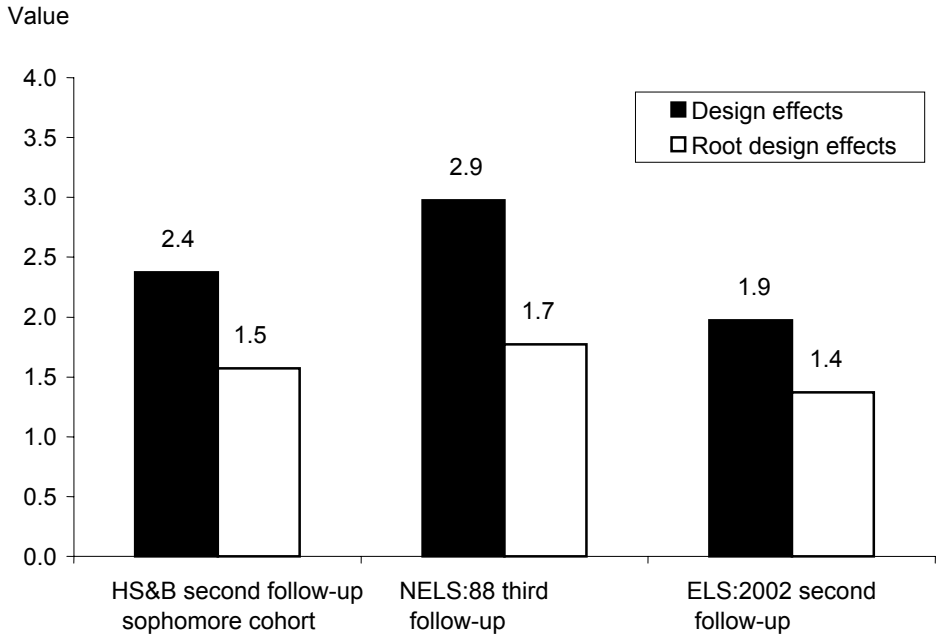
[1] "White and all other races" is predominantly White, with a very small number of individuals from other race categories. All race categories exclude individuals of Hispanic or Latino origin.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Base Year, 2002" and "First Follow-up, 2004."

As discussed in section 3 of this chapter, trimming weights reduces the variance which reduces the design effect. Additionally, the items used to compute the mean design effects were different in the second follow-up than in the base year and first follow-up, because the design effects were not expected to change much between the two rounds of the study. It is more important to compare design effects across cohorts, as described below, so the items were chosen to be as comparable to NELS:88 third follow-up items as possible.

The design effects indicate that the ELS:2002 second follow-up full sample was more efficient than the NELS:88 third follow-up full sample and the HS&B second follow-up sophomore cohort full sample. For means and proportions based on second follow-up questionnaire data for all respondents, the average design effect in ELS:2002 was 1.90; the comparable figures were 2.94 for the NELS:88 third follow-up and 2.40 for the HS&B sophomore cohort second follow-up. Figure 14 shows the mean design effects and root design effects for the HS&B second follow-up sophomore cohort, NELS:88 third follow-up, and ELS:2002 second follow-up.

Figure 14. Full sample mean design effects and root design effects, by longitudinal study: Selected years, 1972–2006



SOURCE: U.S. Department of Education, National Center for Education Statistics, High School and Beyond (HS&B), "Second Follow-up, 1984"; National Education Longitudinal Study of 1988 (NELS:88), "Third Follow-up, 1994"; and Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

The smaller design effects in ELS:2002 compared with those for NELS:88 and HS&B are probably due to subsampling. No subsampling was conducted in the ELS:2002 second follow-up, but additional subsampling was done in the other studies. In NELS:88, subsampling was performed in the first, third, and (not relevant to ELS:2002 comparisons) fourth follow-ups. (See Haggerty et al. [1996] for relevant details.) In HS&B, sophomore cohort members were subsampled for inclusion in the HS&B high school transcript study and this subsample was the basis for the HS&B second follow-up study. (See Zahs et al. [1995] for more details.) The general tendency in longitudinal studies is for design effects to lessen over time, as dispersion reduces the original clustering. However, subsampling increases design effects because it introduces additional variability into the weights with an attendant loss in sample efficiency.

The smaller design effects in ELS:2002 compared with those for the HS&B sophomore cohort also may reflect the somewhat smaller cluster size used in the latter survey in the base year. Although the clusters were reduced somewhat in the first follow-up for both studies, a number of students remained in the base-year school. The HS&B base-year sample design called for 36 sophomores selected from each school. The ELS:2002 sample design called for about 26 sophomores selected from each school. Clustering tends to increase the variance of survey estimates because the observations within a cluster are similar and therefore add less information than independently selected observations. The impact of clustering depends mainly on two factors: the number of observations within each cluster and the degree of within-cluster homogeneity. When cluster sizes vary, the impact of clustering (DEFFc) can be estimated by

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

where $\bar{b}$ refers to the average cluster size (the average number of students selected from each school) and rho refers to the intraclass correlation coefficient, a measure of the degree of within-cluster homogeneity. If the value of rho (which varies from one variable to the next) averaged about 0.05 in both studies, then the reduced cluster size in ELS:2002 would almost exactly account for the reduction in the design effects relative to HS&B.

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

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

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

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

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

Tables 67 and 68 make it clear that the DEFFs and DEFTs vary considerably by subgroup. It is therefore important to use the mean DEFT for the relevant subgroup in calculating approximate standard errors for subgroup statistics.

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

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

$$Var(b-a) = Var(b) + Var(a)$$

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

A final principle is that more complex estimators show smaller design effects than simple estimators (Kish and Frankel 1974/2003). Thus, correlation and regression coefficients tend to have smaller design effects than subgroup comparisons, and subgroup comparisons have smaller design effects than means. This principle implies that it will be conservative to use the DEFTs in the above tables in calculating approximate standard errors for complex statistics, such as multiple regression coefficients. The procedure for calculating such approximate standard errors is the same as with simpler estimates: first, a standard error is calculated using the formula for data from a simple random sample; then the standard error is multiplied by the appropriate DEFT.

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

6.5 Second Follow-up Imputation

6.5.1 Imputation Variables

Five key analysis variables were selected for imputation for the ELS:2002 second follow-up study. These were five new variables from the second follow-up study. Table 70 lists the selected variables. The five variables selected for imputation include indicators of whether the respondent ever applied to or attended a postsecondary institution, whether the respondent ever held a job for pay since high school, total job earnings in 2005 calendar year, and expectations for the highest level of education to be obtained. These variables were chosen because they are classification variables typically used in NCES's descriptive reporting.

Table 70. Second follow-up imputation variables, by number and weighted proportion imputed: 2006

Variable	Number of cases imputed	Weighted percent imputed[1]
Ever attended a postsecondary institution (F2EVRATT)	#	0.00
Ever applied to a postsecondary institution (F2EVRAPP)	#	0.01
Ever held a job for pay since high school (F2EVRJOB)	50	0.41
Highest level of education expected to complete (F2STEXP)	60	0.44
Total job earnings in 2005 calendar year (F2JOBERN)	2,000	14.67

Rounds to zero.

[1] The denominator used in calculating the weighted percent missing varies by variable due to restrictions on eligibility for imputation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

6.5.2 Imputation Methodology

The ELS:2002 second follow-up data were imputed using weighted sequential hot deck imputation (Cox 1980) which was used to impute all five variables. Sequential hot deck imputation is a common procedure used for item nonresponse. This method uses the respondent survey data (donors) to provide imputed values for records with missing values. The basic principle of sequential hot deck imputation involves defining imputation classes, which generally consist of a cross-classification of covariates, and then replacing missing values sequentially from a single pass through the survey data within the imputation classes. When sequential hot deck imputation is performed using the sampling weights of the item respondents and nonrespondents, the procedure is called weighted sequential hot deck imputation. This procedure takes into account the unequal probabilities of selection in the original sample by using the sampling weight to specify the expected number of times a particular respondent's answer was used to replace a missing item. These expected selection frequencies are specified so that, over repeated applications of the algorithm, the expected value of the weighted distribution of the imputed values will equal in expectation within imputation class the weighted distribution of the reported answers.

6.5.3 Imputation Results

Similar to the base-year and first follow-up studies, these key variables were imputed for second follow-up respondents where a respondent is defined as a sample member who completes a sufficient portion of the questionnaire. The order in which variables were imputed depended on whether the response of one variable was dependent on the response of another variable. For example, the variable describing whether the respondent ever attended a postsecondary institution was imputed after the variable describing whether the respondent ever applied to a postsecondary institution. Similarly, the variable describing total job earnings in calendar year 2005 is dependent on the variable describing whether the respondent ever held a job for pay since high school. Within these dependencies, the variables were imputed starting with the variable containing the lowest percent missing up to the variable with the highest percent missing. Table 71 presents the imputation classes and sorting variables used in the weighted sequential hot deck imputation procedure. Table 72 presents the before and after weighted distributions for the imputed variables.

6.5.4 Imputation Evaluation

The key measure for determining whether the imputation methods produce acceptable results is that the before- and after-imputation weighted distributions are similar. For evaluation of the imputation results, distributions were considered to be similar when absolute differences are less than 5 percent where the absolute difference is calculated by subtracting the before-imputation weighted percent from the after-imputation weighted percent. If absolute differences were greater than 5 percent, then the unweighted distributions were examined to see if the large differences were due to small sample sizes. Any large differences were evaluated and corrected when possible (for example, by using different imputation classes), and documented when no resolution was possible.

Table 71. Order of imputation variables and variables used in CHAID analysis: 2006

Imputation variable	Sort variables	Predictor variables
Ever applied to a postsecondary institution (F2EVRAPP)	Geographic region of school (BYREGION) School type (BYSCTRL) School urbanicity (BYURBAN)	Enrollment status (F1ENRFIN) Student race/ethnicity (F1RACE) Highest level of education expected to complete (F1STEXP) Current occupation-coded (F2CURROCC) Highest level of education attempted (F2EDLEVL) Ever dropped out (F2EVERDO) Grade level spring term 2004 (F2F1GRDE) High school completion status in 2006 (F2HSSTAT) Respondent type (F2RTYPE) Student sex (F2SEX)
Ever attended a postsecondary institution (F2EVRATT)	Geographic region of school (BYREGION) School type (BYSCTRL) School urbanicity (BYURBAN)	Enrollment status (F1ENRFIN) Student race/ethnicity (F1RACE) Highest level of education expected to complete (F1STEXP) Current occupation-coded (F2CURROCC) Highest level of education attempted (F2EDLEVL) Ever dropped out (F2EVERDO) Grade level spring term 2004 (F2F1GRDE) High school completion status in 2006 (F2HSSTAT) Respondent type (F2RTYPE) Student sex (F2SEX)
Highest level of education expected to complete (F2STEXP)	Geographic region of school (BYREGION) School type (BYSCTRL) School urbanicity (BYURBAN)	Enrollment status (F1ENRFIN) Student race/ethnicity (F1RACE) Highest level of education expected to complete (F1STEXP) Current occupation-coded (F2CURROCC) Highest level of education attempted (F2EDLEVL) Ever dropped out (F2EVERDO) Grade level spring term 2004 (F2F1GRDE) High school completion status in 2006 (F2HSSTAT) Respondent type (F2RTYPE) Student sex (F2SEX)
Ever held a job for pay since high school (F2EVRJOB)	Geographic region of school (BYREGION) School type (BYSCTRL) School urbanicity (BYURBAN)	Enrollment status (F1ENRFIN) Student race/ethnicity (F1RACE) Highest level of education expected to complete (F1STEXP) Current occupation-coded (F2CURROCC) Highest level of education attempted (F2EDLEVL) Ever dropped out (F2EVERDO) Grade level spring term 2004 (F2F1GRDE) High school completion status in 2006 (F2HSSTAT) Respondent type (F2RTYPE) Student sex (F2SEX)
Total job earnings in calendar year 2005 (F2JOBERN)	Geographic region of school (BYREGION) School type (BYSCTRL) School urbanicity (BYURBAN)	Enrollment status (F1ENRFIN) Student race/ethnicity (F1RACE) Highest level of education expected to complete (F1STEXP) Current occupation-coded (F2CURROCC) Highest level of education attempted (F2EDLEVL) Ever dropped out (F2EVERDO) Grade level spring term 2004 (F2F1GRDE) High school completion status in 2006 (F2HSSTAT) Respondent type (F2RTYPE) Student sex (F2SEX)

NOTE: CHAID = Chi-squared automatic interaction detection analysis.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

Table 72. Weighted distribution of imputed variables before and after imputation: 2006

Variable name	Variable description	Variable category	Before imputation		After imputation	
			Sample size	Weighted percent	Sample size	Weighted percent
F2EVRAPP	Ever applied to postsecondary school	Total	14,100	100.0	14,100	100.0
		Legitimate skip	110	1.0	110	1.0
		No	2,700	21.9	2,700	21.9
		Yes	11,400	77.1	11,400	77.1
F2EVRATT	Ever attended postsecondary school	Total	14,100	100.0	14,100	100.0
		Legitimate skip	110	1.0	110	1.0
		No	3,500	28.9	3,500	28.9
		Yes	10,500	70.1	10,500	70.1
F2STEXP	Highest level of education respondent expects to complete	Total	13,000	100.0	14,100	100.0
		Less than high school graduation	30	0.3	30	0.3
		GED or other equivalency only	200	1.8	200	1.8
		High school graduation only	680	5.7	680	5.7
		Attend or complete 2-year college/school	2,000	16.1	2,000	16.1
		Attend college, 4-year degree incomplete	370	2.8	370	2.8
		Graduate from college	4,500	31.6	4,500	31.6
		Obtain a master's degree or equivalent	3,500	23.2	3,500	23.1
		Obtain Ph.D., M.D., or other advanced degree	1,800	10.8	1,800	10.8
		Don't know	1,100	7.7	1,100	7.7
F2EVRJOB	Ever held a job for pay since leaving high school	Total	14,100	100.0	14,100	100.0
		Legitimate skip	110	1.0	110	1.0
		No	1,200	7.4	1,200	7.4
		Yes	12,800	91.6	12,900	91.6
F2JOBERN	Respondent's total 2005 job earnings	Total	13,800	100.0	14,100	100.0
		Legitimate skip	1,200	7.9	1,200	7.7
		No income	390	2.8	400	2.8
		Less than $1,000	1,000	6.9	1,100	6.9
		$1,000 to $2,999	2,700	18.0	2,700	18.0
		$3,000 to $5,999	2,900	20.8	3,000	20.8
		$6,000 to $9,999	1,900	13.9	2,000	14.0
		$10,000 to $14,999	1,600	12.8	1,700	12.8
		$15,000 to $19,999	900	7.3	900	7.4
		$20,000 to $24,999	500	4.5	600	4.6
		$25,000 to $34,999	410	3.5	430	3.7
		$35,000 to $49,999	120	0.9	120	1.0
		$50,000 and above	60	0.6	70	0.6

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002), "Second Follow-up, 2006."

6.6 Data Security; Second Follow-up Disclosure Risk Analysis and Protections

Data security was a pervasive concern for the second follow-up. Extensive confidentiality and data security procedures were employed for ELS:2002 data collection and data processing

activities; some of those procedures are summarized briefly here. All project staff signed confidentiality agreements and affidavits of nondisclosure and are prohibited by law from using the obtained information for any purposes other than this research study. The second follow-up interview data were collected via the web on a server protected with a Secure Sockets Layer encryption policy, which forces all data transferred to or from the website to be encrypted and transmitted only via secure (HTTPS) connection to conforming web browsers. Sample members received an e-mail and a lead letter that described the purpose of the study, that contained the URL to the ELS:2002 secure website, and a user ID number and strong randomly generated credential which allowed them access to the web-based interview. The only mechanism of access to the self-administered web-based interview was through this ID number and credential. Sample members could only access their individual case using their ID number and credential; they could not access data or information about anyone else. The ID numbers provided to sample members were completely different from the data IDs included on the ECB and DAS. Data were prepared in accordance with NCES-approved disclosure avoidance plans. The data disclosure guidelines are designed to minimize the possibility of a data user being able to identify individuals on the file by matching outliers or other unique data to external data sources.

Because of the paramount importance of protecting the confidentiality of NCES data that contain information about specific individuals, ELS:2002 second follow-up data files were subject to various procedures to minimize disclosure risk. The ELS:2002 second follow-up data products and the disclosure treatment methods employed to produce them are described in the following sections.

6.6.1 Second Follow-up Data Products

The set of data products produced for the ELS:2002 second follow-up are different than the set of data products produced in the base year and first follow-up in that no public-use data file was created for the second follow-up. A restricted-use data file and a file developed for use with the NCES DAS were created.

The disclosure treatment developed for the ELS:2002 second follow-up is composed of several steps:

- Review the collected data and identify items that may increase risk of disclosure.
- Apply disclosure treatment[82] to these risky items in order to lower risk of disclosure.
- Produce a restricted-use data file that incorporates the disclosure treated data.
- Produce a file for the DAS that is derived as a subset of items in the disclosure treated restricted-use data file.

The disclosure treatment methods used to produce the ELS:2002 second follow-up data files include variable recoding, variable suppression, and swapping. These methods are described below.

[82] The NCES Statistical Standards (Seastrom 2003) (http://nces.ed.gov/statprog/2002/std4_2.asp), specifically NCES Standard 4-2, provide information both about the legislative background and legal requirements of maintaining confidentiality, and definitions of key terms (perturbation, coarsening, disclosure risk analysis, data swapping, and so forth).

6.6.2 Recoding, Suppression, and Swapping

Some of the data used during data collection activities were deemed to be too identifying and were not included in the restricted-use data file or the file for the DAS. Some restricted-use data were deemed to be too identifying for inclusion in the file for the DAS and these data were not included in the file for the DAS.

For items in the restricted-use file, recoding was used to produce more analytically useful variables. Some items had values that occurred with extremely low frequencies and the items were therefore recoded in order to ensure that all values of all items occurred with a reasonable frequency. Some items included in the file for the DAS were created by producing a recoded version of a restricted-use item. Since the DAS employs an automatic cell suppression methodology that suppresses cell values if the number of responders providing data for that cell is below a certain threshold, recoding of restricted-use items for inclusion in the DAS was carried out in order to reduce the number of cells that would be suppressed by the DAS, thereby increasing the analytic utility of the data included in the DAS.

Swapping was applied to ELS:2002 data items determined to potentially increase risk of disclosure. Respondents were randomly selected for swapping to achieve a specific, but undisclosed, swapping rate. In data swapping, the values of the variables being swapped are exchanged between carefully selected pairs of records: a target record and a donor record. By so doing, even if a tentative identification of an individual is made, because every case in the file has some undisclosed probability of having been swapped, uncertainty remains about the accuracy and interpretation of the match. The swapping was done independently of the swapping conducted in the base year and first follow-up.

Since perturbation (swapping) of the ELS:2002 data may change the relationships between data items, an extensive data quality check was carried out in order to limit the impact of swapping on relationships. Before-and-after weighted distributions and correlations for swapped variables show that, after applying the disclosure limitation techniques, the analytic utility of the data files was not compromised.

6.7 Second Follow-up Unit and Item Nonresponse Bias Analysis

6.7.1 Unit Nonresponse Bias Analysis

Unit nonresponse causes bias in survey estimates when the outcomes of respondents and nonrespondents are different. For the ELS:2002 second follow-up, student response is defined as the sample member completing at least a specified portion of the questionnaire. The weighted response rate[83] was 84.5 percent overall and was greater than 85 percent for all but one of the 33 domains considered in the nonresponse bias analysis. The domains selected for the unit nonresponse bias analysis were derived from the domains listed in section 6.3. Examples of domains used in the nonresponse bias analysis are given below:

[83] Readers are reminded that a smaller denominator was used for general response rate calculations, based on cases actually fielded, than for response rate calculations for weighting purposes. This is because the unfielded cases must be accommodated in the nonresponse adjustments. These unfielded cases in the response rate denominator include sample members who failed to participate in both the base year and first follow-up, freshened students who did not respond in the first follow-up, and a handful of sample members who asked to withdraw from the study.

- Spring 2002 Black 10th-grade students;
- Spring 2002 Hispanic 10th-grade students;
- Spring 2002 Asian 10th-grade students;
- Spring 2002 White/Other 10th-grade students;
- Spring 2002 Public School 10th-grade students;
- Spring 2002 Catholic School 10th-grade students;
- Spring 2002 Other Private School 10th-grade students;
- Spring 2002 10th-grade students who graduated by August 31, 2004;
- Spring 2004 Black 12th-grade students;
- Spring 2004 Hispanic 12th-grade students;
- Spring 2004 Asian 12th-grade students;
- Spring 2004 White/Other 12th-grade students;
- Spring 2004 Public School 12th-grade students;
- Spring 2004 Catholic School 12th-grade students;
- Spring 2004 Other Private School 12th-grade students; and
- Spring 2004 12th-grade students who graduated by August 31, 2004.

The response rate was below 85 percent for one domain (the racial group White/Other), so a nonresponse bias analysis was conducted for this domain. Since the overall response rate was below 85 percent, nonresponse bias analyses were conducted as required under NCES standards. Cross-sectional and panel weights were used in the nonresponse bias analyses.

The nonresponse bias was estimated for variables known for both respondents and nonrespondents. Since the sample for the second follow-up study consists of respondents from the base-year or first follow-up studies, sample member data were used in the nonresponse bias analysis, though some of the available data may have been imputed.[84] The sample member data that were used include:

- student race/ethnicity;
- student sex;
- student's native language;
- family composition;
- parents' highest level of education;
- mother/female guardian's occupation;

[84] For example, some base-year nonrespondents were sampled for inclusion in the first follow-up study. Some of these base-year nonrespondents responded in the first follow-up and some base-year data were collected on these individuals. If these individuals did not provide these base-year data in the first follow-up questionnaires then some of their base-year data were imputed.

- father/male guardian's occupation;
- total family income from all sources; and
- SES.

The sample member's spring 2004 enrollment status was also used and defined as follows:

- in school, in grade (in grade 12);
- in school, out of grade (in grade 10 or 11, ungraded, or graduated early); and
- out of school (dropout or homeschooled).

We also used the sample member cohort flags:

- G10COHRT—indicates a member of the sophomore cohort (i.e., spring 2002 10th-grader); and
- G12COHRT—indicates a member of the senior cohort (i.e., spring 2004 12th-grader).

There were also extensive data available for schools from the base-year school administrator questionnaire, so these data were used to help reduce potential nonresponse bias. Students were linked to the base-year school from which they were sampled. The school sampling frame constructed from the CCD and PSS also contains data for all base-year schools. School data used included the following:

- school sector;
- urbanicity;
- region;
- sophomore enrollment;
- total enrollment;
- number of minutes per class;
- number of class periods;
- number of school days;
- number of students receiving free or reduced-price lunch;
- number of full-time teachers;
- percentage of full-time teachers certified;
- number of part-time teachers;
- number of different grades taught at the school;
- school level;
- coeducational status;
- percentage of students with an Individualized Education Program;

- percentage of students with limited English proficiency;
- percentage Hispanic or Latino sophomores;
- percentage Asian sophomores;
- percentage Black or African American sophomores; and
- percentage all other race sophomores (includes White).

The procedures used for the nonresponse bias analysis were similar to those used in the base year and first follow-up. First, sample member data known for both respondents and nonrespondents were identified. Second, since the set of data known for both respondents and nonrespondents was limited, all of these data were incorporated into nonresponse models used for the second follow-up. The nonresponse adjustments described in section 6.3 were designed to significantly reduce or eliminate nonresponse bias for variables included in the models. Variables not known for most respondents and nonrespondents could not be included in the nonresponse adjustments, and therefore nonresponse bias could not explicitly be reduced for these variables. However, many of the variables in the nonresponse models are correlated with many of the other variables.

Third, after the sample member weights were computed, remaining bias for data known for most respondents and nonrespondents was estimated and statistically tested to check if there was any remaining significant nonresponse bias. Fourth, the remaining bias after student weight adjustments was divided by the standard error, that is, bias/standard error.

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

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

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

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

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

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

or equivalently

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

This formula shows that the estimate of the nonresponse bias is the difference between the mean for respondents and nonrespondents multiplied by the weighted nonresponse rate. The

variance of the bias was computed using Taylor series estimation in RTI's software package SUDAAN.

Tables K-1 and K-2 in appendix K show the nonresponse bias before and after weight adjustments for selected variables for sample members where F2QWT is used in table K-1 and F2F1WT is used in table K-2. The first set of columns in each table shows the estimated bias before nonresponse adjustment for the variables available for most responding and nonresponding students. Statistical tests (*t* tests) were used to test each level of the variables for significance of the bias at the 0.05/(c-1) significance level, where c is the number of categories (levels) within the primary variable. Below is a summary of the before-adjustment significant bias for tables K-1 and K-2:

- At least one level of 19 of the 33 variables was biased for the cross-sectional weight and 21 of the 33 for the panel weight.
- Thirty-seven levels of variables were found to be significantly biased for the cross-sectional weight and 38 for the panel weight.
- Significant biases were usually small.

The second set of columns in tables K-1 and K-2 shows the estimated bias after weight adjustments (using F2QWT for table K-1 and F2F1WT for table K-2) for the variables available for most responding and nonresponding students. The bias after weight adjustments was computed as the difference between the estimate using nonresponse-adjusted and calibrated (final) weights and the estimate using the design (base) weights prior to nonresponse and calibration adjustment. This latter estimate is an estimate of π because it is the estimate of the target population using the design weights. Similar to the testing of before-adjustment bias, *t* tests were performed to test the significance of the bias for each level of the variables. In both tables K-1 and K-2, the estimated bias usually decreased after weight adjustments. Therefore, the number of significantly biased levels of variables decreased from 37 *before* adjustment to 10 *after* adjustment in table K-1 and from 38 *before* adjustment to 16 *after* adjustment in table K-2. In table K-2, the amount of significant bias increased for three levels in two variables. In table K-3, the amount of significant bias increased for eight levels in five variables.

Tables K-3 and K-4 in appendix K show the nonresponse bias before and after weight adjustments for selected variables in the single domain (White/Other race category) where the response rate was less than 85 percent. F2QWT was used in table K-3 and F2F1WT was used in table K-4. As in tables K-1 and K-2, the first set of columns in each table shows the estimated bias before nonresponse adjustment for the variables available for most responding and nonresponding students. Statistical tests (*t* tests) were again used to test the significance of the bias at the 0.05/(c-1) significance level. Below is a summary of the before-adjustment significant bias for tables K-3 and K-4:

- At least one level of 23 variables and a total of 45 levels were found to be significantly biased in table K-3.
- At least one level of 13 variables and a total of 24 levels were found to be significantly biased in table K-4.
- Significant biases were usually small.

As in tables K-1 and K-2, the second set of columns in tables K-3 and K-4 shows the estimated bias after weight adjustments (using F2QWT for table K-3 and F2F1WT for table K-4) for the variables available for most responding and nonresponding students. The bias after weight adjustments was computed the same way as described above for tables K-1 and K-2. Statistical tests (*t* tests) were performed to test the significance of the bias for each level of the variables. In both tables, the estimated bias sometimes decreased after weight adjustments and sometimes increased after weight adjustments. In tables K-3 and K-4, the amount of significant bias increased for four levels among three variables and for five levels among two variables, respectively. Note that sample members are assigned to these domains based on data known for respondents and nonrespondents, and sample members may actually be in different domains. Therefore, these bias estimates are approximate.

The nonresponse bias analyses in conjunction with the weighting adjustments described above do not eliminate all bias. They reduce bias for some of the variables known for most respondents and nonrespondents, which are considered to be some of the analytically important variables and are correlated with many of the other variables. Significant bias after weight adjustments is minimal for the variables analyzed. Some of these variables are used to help create composite (or derived) variables. There may be bias remaining in other variables.

Figures K-1 through K-4 in appendix K compare the estimated relative bias before nonresponse and calibration adjustment with the estimated relative bias after nonresponse and calibration adjustment. Figures K-1 and K-2 examine relative bias for the entire ELS:2002 second follow-up sample using F2QWT and F2F1WT, respectively. Figures K-3 and K-4 examine relative bias for the single domain (White/Other race category) identified as having less than an 85 percent response rate with figure K-3 using F2QWT and figure K-4 using F2F1WT. Relative bias is the bias of the estimate divided by the estimate. It provides an indication of the order of magnitude of the bias with respect to the estimate. Figures K-1 through K-4 indicate that when the relative bias was large before nonresponse adjustment, it was almost always reduced after nonresponse adjustment. When the relative bias was small before nonresponse adjustment, it stayed small after nonresponse adjustment with occasional small increases. These figures clearly show that the nonresponse adjustment reduced bias for sample members.

Nonresponse bias can have an effect on significance testing. Tables K-1 through K-4 include an estimate of the bias ratio (sample bias divided by the standard error). If this ratio is larger than 2 percent, then the probability of a Type I error is greater than 0.05. Figures K-5 through K-8 in appendix K show the sample bias ratio by the Type I error rate. Figures K-5 and K-6 examine bias ratios for the entire ELS:2002 second follow-up sample using F2QWT and F2F1WT, respectively. Figures K-7 and K-8 examine the bias ratios for the single domain (White/Other race category) identified as having a response rate less than 85 percent. F2QWT is used in figure K-7 and F2F1WT is used in figure K-8. Figure K-5 shows that for many of the sample member variables included in the nonresponse bias analysis, the Type I error rate is at or is close to 0.05, and outliers were not graphed. Figures K-6 through K-8 show that although some variables have a Type I error rate at or near 0.05, there are more variables that have a higher Type I error rate. These figures do not take the school bias ratio into account. The school bias ratio varies by school variable, as shown in the ELS:2002 base-year data file user's manual (Ingels et al. 2004). If it is assumed that the school bias ratio is zero, then there is no effect on the sample member bias ratio. However, if the school bias ratio is large, then the Type I error rates

are larger. Although the tables above show that nonresponse bias is minimal, the data user should exercise caution when conducting statistical tests.

6.7.2 Item Nonresponse Bias Analysis

Since the overall weighted unit response rate (84.5 percent) was less than 85 percent, an item nonresponse bias analysis was carried out as required under NCES statistical standards. The first step in the nonresponse bias analysis was to calculate the weighted[85] response rate for every questionnaire item included in the ELS:2002 second follow-up. Four items were found to have response rates lower than 85 percent:

- Date of marriage.[86] (F2D02P/F2D02R)
- Which of the following are reasons why you decided not to continue your education right after high school? (F2B11NA)
- Which of the following are reasons why you have not continued your education after high school? (F2B08NA)
- How did you earn the GED or equivalency, or in other words, what program or school were you enrolled in, if any? (F2A04A)

These items had weighted response rates of 82.9, 61.2, 58.4, and 36.7 percent, respectively. Tables K-5 through K-8 compare item respondents and nonrespondents to these four items using six characteristics known for more respondents and nonrespondents. Weighted distributions of the values of these six characteristics were generated using both respondents and nonrespondents, using respondents only, and using nonrespondents only and these distributions are presented in tables K-5 through K-8. It should be noted that all unweighted sample counts were rounded for reporting purposes.

Three statistically significant biases (table K-5) were identified for the item Date of Marriage. No statistically significant biases (table K-6) were identified for the item F2B11NA. Two statistically significant biases (table K-7) were identified for the item F2B08NA. One statistically significant bias (table K-8) was identified for the item F2A04A.

Six of the 76 bias comparisons yielded a statistically significant bias. Four of the six statistically significant biases indicate overrepresentation of females or Whites/other among the respondents, as compared to the nonrespondents, and are the largest biases among all six.

[85] Weighted response rates were calculated using the F2 cross-sectional weight, F2QWT.
[86] The restricted-use version of this variable gives month and year while the public-use version only gives quarter and year.

Chapter 7
Data File Contents

This chapter describes the Education Longitudinal Study of 2002 (ELS:2002) base-year to first follow-up and base-year to second follow-up longitudinal data file contents. It addresses the following topics: the structure of the electronic codebook (ECB) system (appendix B), including the megafiles; the nature of the Data Analysis System (DAS); and the questionnaire and composite variables, including their naming conventions and an overview of composite variables (also see appendix L, ECB and DAS variable list; appendix M, list of composite variables; and appendix N, variables imported into ELS:2002 from external sources).

7.1 Base-Year to First Follow-up ECB Data Structure

ELS:2002 base-year to first follow-up data have been made available in public- and (for licensed users) restricted-use versions[87] in an ECB format on CD-ROM. The ECB is designed to be run in a Microsoft Windows environment. A version of the restricted ECB with high school transcript data added was released in November 2006. (This version is called E4T [NCES 2006-351]; however, the transcript data are also included on the second follow-up [2006] restricted release.) At the same time that the transcript and course offerings data were added in, a final first follow-up test score was added as well. This was the concordance score linking the scales of the 2005 National Assessment of Educational Progress mathematics assessment to the 2004 ELS:2002 math score; the concordant scale score is described at length in chapter 2 of this volume.

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

The base-year to first follow-up ECB comprises two large "megafiles," one at the student level (with other data sources supplying contextual data for analysis of the student) and one at the high school level. The megafile at the student level encompasses base-year student (student questionnaire and test, parent, and teacher questionnaires) and school (administrator, library, facilities) data in conjunction with first follow-up student (student, transfer, dropout, early graduate, and homeschool questionnaires, student tests and transcripts) and school administrator data.

The second megafile, at the school level, encompasses base-year data (facilities checklist, the school administrator questionnaire, the library media center questionnaire) and first follow-

[87] A license is required to access the restricted-use ECB (http://nces.ed.gov/pubsearch/licenses.asp).

up school administrator questionnaire and course offerings data. Analysts should be aware that the base-year school data may be used as a standalone, nationally representative sample of 2001–02 schools with 10th grades, but that the school data for the 2003–04 school year are *not* precisely generalizable to the nation's 2003–04 high schools with 12th grades.

The content and organization of the transcript and course offerings data (course-level file, student-level file, school-level file, and course offerings file) are further described in Bozick et al. (2006).

7.1.1 Base-Year to Second Follow-up ECB Data Structure

The base-year to second follow-up data are available in a restricted-use ECB (NCES 2008-346) on CD-ROM. This ECB contains all of the base-year to first follow-up data (including high school transcript data) as well as the second follow-up data. The structure of the new base-year to second follow-up ECB builds on the past ECBs but contains additional dimensions. Again, there are both student and high school-level megafiles but there is also a postsecondary institutional file and an extant data sources file that reflects ancillary data imported from external administrative records. A "Quick Guide" for using the ECB is included in this report as appendix B.

7.1.2 Student Megafile

The student file contains all prior-round data,[88] retaining the basic structure as in the base-year to first follow-up Transcript ECB (E4T: NCES 2006-351). New variables were usually added to new sections and then inserted into a logical grouping of sections (i.e., composites, sample member response data, school replicated data, etc.). The section titled "ID and Universe Variables" is an exception in spanning rounds of data collection.

Sections of the student file (BYF2STU) are as follows:

- ID and Universe Variables;
- Base-year (BY) Weights and Composites;
- First follow-up (F1) Weights and Composites;
- F1 Transcript Composites;
- Second follow-up (F2) Weights and Composites;
- Second follow-up Extant Data Source Composites;
- BY Student Questionnaire;
- F1 Student Questionnaire;
- F1 Dropout Questionnaire;
- F1 Transfer Questionnaire;
- F1 Early Graduate Questionnaire;

[88] While all data elements have been retained, not all base-year and first follow-up data have been carried over. Specifically, in two rare instances data have been expunged: past data for deceased sample members, and data for sample members who withdrew their participation with instructions that past data be dropped.

- F1 New Participant Supplement;
- F2 Survey;
- BY Parent Questionnaire;
- BY Teacher Questionnaire (English);
- BY Teacher Questionnaire (Math);
- BY School Composites;
- F1 School Composites;
- BY Administrator Questionnaire;
- F1 Administrator Questionnaire;
- BY Library Questionnaire; and
- BY Facilities Checklist.

7.1.3 High School Megafile

The school file reflects data for the base-year, first follow-up, and first follow-up transcript data collection; the first follow-up was the final round for collection of school-level data directly from high schools. Common Core of Data and Private School Survey data were added to the restricted-use ECB as a convenience to the ECB user. The School ID is constructed such that student file records can be merged with the high school data.

7.1.4 Postsecondary Institution File

The postsecondary institution file is newly added with the second follow-up and links students to postsecondary institutions applied to and attended. The key on the file is Stu_ID, order number, and Integrated Postsecondary Education Data System ID. Data for the institutions are obtained in the second follow-up interview, and collected by looping over each institution for a series of questions about application and attendance, among others. The looped iterations were normalized (one record for each unique postsecondary institution per caseid) and placed into the institution file structure. The order number enables researchers to associate information for a given institution from the student-level file with information about the given institution. An order number helps researchers determine a uniquely identifiable key and to allow users to easily link institution-based items from the student file to the institution file.

If the respondent reported attending one postsecondary institution, this institution is listed first for that student. If the respondent indicated attending more than one postsecondary institution, the one the respondent attended first would be listed first and so on. Institutions that respondents applied to but did not attend follow in the order they were named in the interview.

7.1.5 Extant Data Source Files: Ancillary Data Links in the ELS:2002 Base-Year to Second Follow-up ECB

Rather than merge data from extant data sources on the student file, separate files were constructed that can be linked to the student file. Sample members will have one record on each data source file when data are available. If information is not available for that data source, then

the student record will be excluded from that data source file. The following data source files were utilized:

- the Central Processing System[89];
- the National Student Loan Data System[90];
- the Scholastic Aptitude Test (SAT);
- the ACT; and
- the General Educational Development (GED).

Variables representing the extant sources data imported into the second follow-up are listed in appendix N of this document. Some composite variables have been constructed to facilitate use of the SAT and ACT test score data. Further details on merged SAT/ACT data may be found later in this chapter (section 7.2.2.3).

7.1.6 Reserve Codes

There are a number of reasons for data to be missing for given variables. We account for these situations by filling items with reserve codes. The following reserve code scheme was used:

- -1 "Don't know. " This reserve code was not used in the second follow-up and is retained for prior-round data.
- -2, "Refused." This reserve code was not used in the second follow-up and is retained for prior-round data.
- -3 "Item legitimate skip/NA." Filled for questions that are not answered because prior answers route the respondent elsewhere.
- -4 "Nonrespondent." Filled for all variables across the entire instrument when a sample member did not respond to the instrument.
- -5 "Out of Range." This reserve code was not used in the second follow-up and is retained for prior-round data.
- -6 "Multiple Response." This reserve code was not used in the second follow-up and is retained for prior-round data.
- -7 "Partial interview-breakoff." Filled for questions that are not answered because the respondent has broken off the interview without completing it. This also includes particular items that were not included on abbreviated versions of previous-round questionnaires.
- -8 "Survey component legitimate skip/NA." Filled for all variables across the entire instrument when a sample member does not apply to a particular instrument or round. It is similar to -4 in that it applies to all variables across an entire instrument;

[89] The Central Processing System contains Free Application for Federal Student Aid data.
[90] The National Student Loan Data System (NSLDS) database contains records of all federal loans, and Pell grant information, for anyone who has such a loan or grant.

however, the reason is different in that the sample member never had the chance to respond.

- -9 "Missing." Filled for questions that are not answered when the routing suggests that they should have responded.

7.1.7 Data Analysis System

In addition to the ECBs, for users who do not require direct access to microdata, ELS:2002 data are also available from the National Center for Education Statistics through a web-based DAS which includes data through 2006 and selected transcript variables (e.g., coursetaking summaries and categorical data for grade point average) from the high school transcript file. The DAS software makes it possible for users to specify and generate their own tables. In addition to the table estimates, the DAS calculates standard errors and weighted sample sizes for these estimates. Finally, the DAS will also produce a correlation matrix of selected variables to be used for linear regression models. Included in the output with the correlation matrix are the design effects for each variable in the matrix. Since statistical procedures generally compute regression coefficients based on simple random sample assumptions, the standard errors must be adjusted with the design effects to take into account the stratified sampling method used in the ELS:2002 surveys. The DAS can be accessed electronically at http://nces.ed.gov/DAS.

The DAS will give essentially, but not precisely, the same estimates and standard errors as the ECB. Because of its rounding conventions, DAS estimates will differ from ECB estimates by being slightly less precise. Because a different method is used for variance estimation, standard errors of measurement, while highly similar, will seldom be identical (the ECB estimates sampling errors through a Taylor Series linearization; the DAS estimates standard errors using the balanced repeated replication method of approximating the estimator by balanced repeated replication of the sampled population).

7.2 Instrument and Composite Variables

7.2.1 Naming Conventions

Data users should find naming conventions for variables, flags, and weights intuitive and quite similar to those employed in the National Education Longitudinal Study of 1988. Most variables begin with an indicator of the wave (e.g., base-year variables begin with BY, first follow-up with F1, and second follow-up with F2). Weights follow the same wave-naming convention and also contain the suffix WT (e.g., BYSTUWT is the name for the final student weight for base-year questionnaire completion, F2QWT is the equivalent second follow-up questionnaire completion weight, and BYSCHWT is the name for the base-year final school weight). Just as first follow-up variables begin with the prefix F1, second follow-up (2006) variables begin with F2.

In the base year and first follow-up (but not the second follow-up), variable names also distinguish (in their third character) between components and questionnaire types. F1S, for example, indicates a first follow-up student questionnaire variable, whereas F1A stands for administrator questionnaire items, and F1D refers to the "out of school" (dropout) questionnaire. Variables that reflect specific items in the questionnaire carry the question number in the variable

name, immediately after the component indicator. Hence, F1S58 would be item 58 from the first follow-up student questionnaire, and F1D19 would be item 19 in the dropout instrument.

The round-specific constructed variables are typically not anchored in a single questionnaire item and may sometimes reflect nonquestionnaire sources of information, such as the assessments. First follow-up test scores carry the prefix F1TX. F1TXMQU, for example, indicates the quartile score for the first follow-up mathematics test. Flags are indicated by the suffix FLG or FG. Variable names also distinguish between the public (P) and restricted (R) use forms, where variables differ between them (the base-year and first follow-up public-use variables are a subset of the restricted-use superset).

Finally, some slightly different information is included in second follow-up variable names. In base year and first follow-up, variable names contain a letter to reference a questionnaire (e.g., S = Student) in addition to the round prefix (BY or F1) and frequently reference the question number (composite and transcript variables do not link to specific questionnaire items so they contain a descriptive reference). The second follow-up instrument is an electronic questionnaire with many pathways; there is no fixed hardcopy questionnaire nor question numbers. However, a sequential number within each thematic area or module has been assigned to each item from the interview. Whenever possible, second follow-up variable names were constructed as F2{Section Letter}{Sequential Number}{sub-item letter if applicable}. The applicable section letters for the 2006 round are as follows:

- A—High school section;
- B—Postsecondary section;
- C—Employment section; and
- D—Community section.

Variables that do not follow the sequential numbering naming convention are:

- Postsecondary institution variables—These variables were obtained at the respondent level and looped through each institution. The final file is normalized with each record representing one of the institutions the respondent identified in the interview. The variables are named with a descriptive reference.
- Composites—These variables were given names consistent with the descriptive names of prior round composites, prefixed with the 2006 round indicator (i.e., F2).

For the ELS:2002 second follow-up, no hardcopy codebooks were produced. For base-year to first follow-up data, the hardcopy codebooks appear as portable document format (PDF) files for the web-published version of the data documentation manual (see http://nces.ed.gov/surveys/els2002) and correspond to appendix G of Ingels et al. (2005b). The codebook supplies a comprehensive description of the student data file. For each variable on the student component data file, the codebook provides a summary of the related information, including the question number and wording, the variable name, and the responses to the item, along with their unweighted frequency and percent and weighted percent. It also provides missing data frequencies sorted by reserve codes. For the high school transcript data, hardcopy codebooks are also available, as an appendix to Bozick et al. (2006). Unlike the other hardcopy codebooks, however, the hardcopy transcript codebooks are only available as part of the restricted-use data.

7.2.2 Second Follow-up Composite Variables

The second follow-up data file includes many composite variables for the convenience of data users. Appendix M provides a complete list of second follow-up composite variables. Composite variables combine or reorganize data whereas instrument variables (that is, variables named with an "F2A," "F2B," "F2C" or "F2D" prefix) represent the data as they were collected in the interview. This section provides a descriptive overview of two types of composite variables. Composite variables that are constructed from multiple data sources will be discussed first. The month-by-month enrollment and employment history composite variables will be covered second. More detailed descriptions of the construction methods used for each of these composite variables and the associated code are provided in the ECB.

7.2.2.1 Composite Variables Constructed From Multiple Data Sources

First, we will provide an overview of the second follow-up composite variables that merge data from multiple sources. In each of these composite variables, data collected from the second follow-up interview is one input. The second follow-up data collection began at the end of January 2006 and continued through early September 2006. Respondents provided information based on their status at the time of their interview.

Many of these composite variables use information collected from the *High School* section of the interview as one input. Owing to the complexity of these variables, they will be treated first in their own subsection. A discussion of composite variables that integrate data from the *Postsecondary Education* section of the second follow-up interview with Integrated Postsecondary Education Data System (IPEDS) data follows.

Composite variables using high school completion data from multiple sources. Many of the second follow-up composite variables draw upon data from the High School section of the interview. The data user is cautioned that most of the variables that provide data as they were collected in the High School section of the second follow-up interview, that is, variables with an "F2A" prefix, are not standalone variables to be used in analyses, but rather they serve as inputs to composite variables. They are provided on the ECB to reflect the direct responses to items administered in this section of the interview and for reference or validation of composite variable construction. They are not included on the DAS.

The *High School* section data are supplemented by data from three primary sources; the first follow-up early graduate and dropout interviews, the high school transcript data as provided on the high school transcript ECB, and the high school transcript data as preloaded in the second follow-up interview. A distinction is drawn between the high school transcript data as provided on the ECB and the preloaded transcript information because these data were still undergoing quality control procedures when the second follow-up data collection began. In an effort to preload only stable transcript data, transcript information was only preloaded for cases where the following conditions were met: (1) the data indicated that a high school diploma or certificate of attendance had been awarded; (2) the high school completion date was May or June 2004, the modal dates of completion; and (3) quality control had been completed. High school completion information as reported in the first follow-up early graduate and dropout questionnaires was also preloaded. F2PHSDG indicates the credential earned as it was preloaded. The preloaded high school completion dates are found in F2PHSDT.

Inevitably, data collected from multiple sources are inconsistent for a small number of cases. Therefore, for the purpose of constructing many composite variables using data from multiple sources, decisions must be made with respect to which data sources take precedence over other data sources.[91] Some of these decisions were "built in" to the preloaded data. Specifically, if the sample member reported earning a high school credential and a completion date in his or her first follow-up early graduate or dropout questionnaire, this information was preloaded instead of any high school transcript data that may exist for that sample member. In other words, high school completion information collected in the first follow-up early graduate or dropout questionnaire was given precedence over high school transcript data. This approach was taken for two reasons. First, as previously mentioned, the high school transcript data were still undergoing quality control procedures for some cases when data collection began. In addition, since the preloaded information was presented to the second follow-up respondent in the interview by way of customized question wording, consistency with the respondents' own perception of their high school completion status was desired. In cases where the preload variables were not populated, second follow-up respondents were only asked if they had completed high school, for the credential they had earned, and when they received that credential if these data were not preloaded. Finally, if data were not available from the preloads or second follow-up responses, the high school transcript data, as provided on the high school transcript ECB, were referenced for some composite variables.

In summary, the precedence order of data sources for composite variables constructed from the *High School* section data is as follows:

1. First follow-up respondent report in the early graduate or dropout questionnaire questionnaires (including but not limited to preloaded high school completion data);
2. Preloaded high school transcript data (high school diploma or certificate of attendance in May or June 2004; see F2PHSDG and F2PHSDT);
3. Second follow-up respondent report in the *High School* section of the interview (F2A variables; only populated if 1 and 2 are not); and,
4. Final high school transcript data (as necessary for some composite variables).

Not all of the high school composite variables draw on all of these data sources however. For example, composite variables that pertain only to second follow-up respondents do not integrate the fourth source, final high school transcript data (except in the rare instances of second follow-up item nonresponse).

A number of these multisource composite variables reference the spring term of 2004, the reference period for the first follow-up data collection. The first of these, F2F1GRDE, updates F1GRADE for first follow-up nonrespondents. It indicates the grade level in the spring term of 2004 for sample members who were attending high school at that time. First follow-up nonrespondents were identified in the second follow-up as spring-term 2004 12th-graders by their response to a direct question about their grade level during that time (F2A12) or by logical imputation based on having received a diploma or certificate of attendance in April, May, or June 2004. In keeping with the classification rules used for first follow-up respondents, first

[91] Some composite variables constructed from multiple data sources incorporate all information from all data sources. For example, the variable F2EVERDO is set to "1" if any source indicates a dropout episode. However, for most composite variables, inconsistencies are reconciled based on decision rules.

follow-up nonrespondents who indicated that they had completed their high school credential prior to April 2004[92] (early graduates) were not included in the 12th-grade cohort. A closely related variable, G12COHRT, indicates which ELS:2002 sample members were in the 12th grade, the modal grade level, in the spring term of 2004.

Another variable that relates to this time period, F2SP04DO, indicates whether the sample member was a spring-term 2004 dropout or early alternative completer as defined by the classification rules used in the first follow-up data collection. A sample member was considered a spring-term 2004 dropout if he or she had experienced a dropout episode of at least 4 consecutive weeks during that term. The dropout episode could have begun prior to the start of the spring term. F2SP04DO identifies a sizable number of first follow-up nonrespondents who were spring-term 2004 dropouts. This variable also identifies a small number of first follow-up respondents who were high school students at the time of their first follow-up interview, but experienced a dropout episode during the spring term of 2004 subsequent to their first follow-up participation. Sample members who completed high school early by earning a GED are also accounted for in this variable. Given its comprehensive nature, this variable may be used for national estimates of dropout status during the term when most cohort members were completing high school.

The composite variables F2WYLV1–F2WYLV14 are populated for the spring-term 2004 dropouts and early alternative completers who are identified in F2SP04DO. These variables indicate the respondents' reasons for dropping out of high school prior to or during the spring term of 2004. These composite variables combine the responses provided in the first follow-up dropout and early graduate questionnaires with the responses provided in the *High School* section of the second follow-up interview. These questions were never asked of the same sample member in both the first and second follow-up interviews. Therefore, inconsistent information from these two sources was not an issue.

High school dropouts as of the second follow-up interview in 2006 are identified in F2HSSTAT. This variable also indicates whether these dropouts reported working toward a GED. The variable is populated for the sample universe: it includes information about both second follow-up respondents and non-respondents.

There are two other variables that identify high school dropouts: F2EVERDO and F2DOSTAT. Unlike F2SP04DO and F2HSSTAT which identify dropouts at a particular period of time, these variables identify individuals who had dropped out of school at any one of the data collection points. The data collection points are the first follow-up, the high school transcript, and the second follow-up data collections and the enrollment status updates between data collections. The enrollment status updates did not reference the entire period of time between data collections. In other words, the information on dropout episodes held in these variables is not comprehensive. A dropout episode which began and ended between any two data collection points would not be detected. F2EVERDO simply indicates whether a dropout episode was detected for a given sample member. F2DOSTAT indicates whether there is any evidence of a dropout episode as well as high school completion status as of the second follow-up interview.

[92] In the first follow-up, early graduates were defined as respondents who completed a high school credential on or before March 15, 2004. Since first follow-up nonrespondents were often completing the second follow-up interview 2 years after earning their high school credential, they were unlikely to remember the precise date of that event. Therefore, only month and year of high school completion were collected in the second follow-up interview. Consequently, early alternative completers identified in the second follow-up are defined as those who earned their GED prior to April 2004.

There are several other variables that relate to dropouts and/or GED recipients. F2HSLVDP (and F2HSLVDR on the ECB) indicates when GED recipients and 2006 dropouts last attended high school. F2GEDPRG (and F2GEDOTH on the ECB) indicates the program through which the GED was earned. F2GEDST indicates the state in which the GED was earned. Reasons for completing a GED are provided in F2WYGED1 through F2WYGED6. All of these variables combine data collected in the first follow-up early graduate and dropout questionnaires with data collected in the *High School* section of the second follow-up interview. Because sample members were never asked to answer these questions twice, the possibility of inconsistent responses from the first and second follow-up interview was prevented. F2GEDPRG, F2GEDOTH, F2GEDST and F2WYGED1-6 are only populated for sample members who reported in their first or second follow-up interview that they had earned a GED. On the other hand, F2EVRGED identifies sample members for whom we have evidence of GED completion from any one of the following sources: first follow-up interview, high school transcript, second follow-up interview and/or data from the American Council on Education (ACE).

Several of the multisource composite variables are related to educational attainment. F2HSSTAT indicates high school completion status as of the second follow-up interview. For those who had completed high school, F2HSCPDP (and F2HSCPDR in the ECB) indicates the high school completion date. F2EDLEVL indicates educational attainment including any postsecondary attendance as reported in the *Postsecondary Education* section of the interview.

F2RTYPE categorizes second follow-up respondents into one of six categories based on the timing of any postsecondary enrollment in relation to their high school completion/exit date; standard enrollee, delayer, leaver, delayer-leaver, nonenrollee and high school student. Eligibility for certain portions of the interview is dependent on the respondent's type (see section 2.5).

Composite variables integrating second follow-up postsecondary education data with IPEDS data. Second follow-up respondents were asked to name the postsecondary institutions to which they had applied (when they first submitted applications), the institutions where they were admitted, and the institution(s) they had attended. The name and location of each institution as entered into the web interview was matched against a list of postsecondary institutions from IPEDS. The correct match was selected from a display of potential matches. When a selection was made, the institution's IPEDS unit ID was stored in the ELS:2002 database. The IPEDS data include a wealth of information on postsecondary institutions. A few key characteristics of these institutions such as state, level of offering (i.e., 4 or more years; at least 2, but less than 4 years; less than 2 years), institutional control (i.e., public, private not-for-profit, private for-profit) and sector (e.g., public, 4-year or above; private not-for-profit, 4-year or above) are included in the ELS second follow-up institution data file for convenience (see F2ISTATE, F2ILEVEL, F2ICNTRL, and F2ISECTR). The level of offering, institutional control, and sector of the first postsecondary institution attended (see F2PS1 and section 2.5) are provided on the sample member file (see F2PS1LVL, F2PS1CTR, F2PS1SEC). In a small number of cases, the data in these composite variables were provided by the respondent rather than the IPEDS data. When a match was not found in the IPEDS institution listing, respondents were asked to provide the state, level of offering, and institutional control of the institution. Project staff later attempted to select the appropriate institution from the IPEDS listing. When successful, the IPEDS information was preferred over the respondent's report of these data. However, if project staff were not able to identify an IPEDS institution for these cases, the information as reported by the

respondent was provided in the composite variables. The ELS:2002 second follow-up institution file also includes the IPEDS unit ID so analysts may link to the IPEDS data to draw upon other variables of interest (see F2IIPED).

7.2.2.2 Month-by-Month Enrollment and Employment History Composites

In the *Postsecondary Education* section of the second follow-up interview, all respondents who reported attending a postsecondary institution since high school were asked to provide the months they were enrolled. The month-by-month enrollment at each institution attended, beginning with January 2004, is provided on the postsecondary institution file (see F2I0401–F2I0608). Since some respondents last attended high school prior to 2004, F2IPRE4 indicates the number of months of postsecondary enrollment since high school in 2002 and 2003. Since some respondents had attended more than one postsecondary institution, a series of composite variables was created to indicate enrollment across institutions (see F2PSPRE4, F2PS0401–F2PS0608). Enrollment at any postsecondary institution in a given month is represented as attendance in these composite variables.

There is also a series of composite variables indicating month-by-month employment status (F2EM0206–F2EM0608). These composites were built from a series of questions posed to second follow-up respondents who indicated that they had not attended a postsecondary institution and were not currently enrolled in high school. Nonenrollees provided, as applicable, the date they began their first job after high school, the date they left that job, and the date they started their current job. Employment was assumed to be continuous between the start and end date of a job. The months for which employment could not be logically imputed based on these dates were referenced in follow-up questions about employment and labor force status. If the respondent indicated employment during one of these months, the composite variable indicates employment. To determine labor force status, respondents who indicated that they were not working in a given month were asked if they were looking for work at that time. The number and percent of months unemployed since high school completion or exit (or since June 2002 if last attended prior to that date) are also provided (F2NUNEMP and F2PUNEMP).

7.2.2.3 Composite Variables Constructed From Transcript and External Data Sources: Blended Test Scores/ACT-SAT Concordance

SAT and ACT test scores were obtained from high school transcripts collected in the ELS:2002 first follow-up in 2005, and from the College Board, and ACT in 2007. These data sources were combined to provide maximum coverage of the subset of the second follow-up sample that had taken either or both of the exams. A concordance between ACT and SAT scores was generated also. If the data source is an ACT score and the composite score is provided in terms of an SAT score, ACT to SAT concordance rules are applied. If the source is SAT data and the composite score is provided in terms of an ACT score, SAT to ACT concordance rules were applied. Concordance rules are explained in the following document: http://www.collegeboard.com/prod_downloads/highered/ra/sat/satACT_concordance.pdf.

The following SAT and ACT scores are available on the ECB:

- TXEESATC—Highest entrance exam composite score (in terms of SAT score);
- TXEEACTC—Highest entrance exam composite score (in terms of ACT score);
- TXEESATM—Highest entrance exam Math score (in terms of SAT score);
- TXACTC—Highest ACT composite score;

- TXACTM—Highest ACT Math score;
- TXACTR—Highest ACT Reading score;
- TXACTE—Highest ACT English score;
- TXACTS—Highest ACT Science score;
- TXSATM—Highest SAT Math score;
- TXSATV—Highest SAT Verbal score; and
- TXSATC—Highest SAT composite score.

AP exam score composites. AP exam scores were obtained from High School transcripts and College Board. A combination of these data sources is used to provide a score for the test score composites. Composites are available for each AP examination subject.

SAT subject test score composites. SAT subject test (SAT II) scores were obtained from high school transcripts and College Board. A combination of these data sources is used to provide a score for the test score composites. Composites are available for each subject test.

Obtaining ECB or DAS. Information on obtaining the restricted-use ELS:2002/06 base-year to second follow-up ECB—as well as information on obtaining the base-year to second follow-up DAS—can be found by reviewing the data products for the study at http://nces.ed.gov/pubsearch. IES/NCES will only accept restricted-use data license applications through its electronic application system (see http://nces.ed.gov/statprog/instruct.asp). More information about applying for restricted-use data licenses is available at http://nces.ed.gov/statprog/instruct.asp and in the *Restricted-Use Data Procedures Manual* at http://nces.ed.gov/statprog/rudman/toc.asp.

References

Angoff, W.H. (1982). Summary and Derivation of Equating Methods Used at ETS. In P.W. Holland and D.B. Rubin (Eds.), *Test Equating* (pp. 55–79). New York: Academic Press.

Beaton, A.E., and Gonzalez, E. (1995). *The NAEP Primer*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Bozick, R., and Ingels, S.J. (2007). *Mathematics Coursetaking and Achievement at the End of High School: Evidence from the Education Longitudinal Study of 2002 (ELS:2002)*. (NCES 2008-306). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Bozick, R., and Lauff, E. (2007). *A First Look at the Initial Postsecondary Experiences of the Sophomore Class of 2002 (ELS:2002)* (NCES 2008-308). U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Bozick, R., Lytle, T., Siegel, P.H., Ingels, S.J., Rogers, J.E., Lauff, E., and Planty, M. (2006). *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation* (NCES 2006-338). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Burns, L.J., Heuer, R., Ingels, S.J., Pollack, J.M., Pratt, D.J., Rock, D., Rogers, J., Scott, L.A., Siegel, P., and Stutts, E. (2003). *ELS:2002 Base Year Field Test Report* (NCES 2003-03). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

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

Carlson, B.L., Johnson, A.E., and Cohen, S.B. (1993). An Evaluation of the Use of Personal Computers for Variance Estimation with Complex Survey Data. *Journal of Official Statistics*, *9*(4): 795–814.

Chromy, J.R. (1981). Variance Estimators for a Sequential Sample Selection Procedure. In D. Krewski, R. Platek, and J.N.K. Rao (Eds.), *Current Topics in Survey Sampling* (pp. 329–347). New York: Academic Press.

Cohen, S.B. (1997). An Evaluation of Alternative PC-Based Software Packages Developed for the Analysis of Complex Survey Data. *The American Statistician*, *57*(13): 285–292.

Cox, B.G. (1980). The Weighted Sequential Hot Deck Imputation Procedure. *Proceedings of the Section on Survey Research Methods* (pp. 721–726). The American Statistical Association.

Curtin, T.R., Ingels, S.J., Wu, S., and Heuer, R. (2002). *NELS:88 Base Year to Fourth Follow-up Data File User's Manual* (NCES 2002-323). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Dalton, B., Ingels, S.J., Downing, J., and Bozick, R. (2007). *Advanced Mathematics and Science Coursetaking in the Spring High School Senior Classes of 1982, 1992, and 2004* (NCES 2007-312). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Deville, J.C., and Särndal, C-E. (1992). Calibration Estimating in Survey Sampling. *Journal of the American Statistical Association*, *87*: 376–382.

Dorans, N.J. (2004). Equating, Concordance and Expectation. *Applied Psychological Measurement*, 28(4): 227–246.

Embretson, S.E., and Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.

Fleiss, J.L., Levin, B.A., and Paik, M.C. (2003). *Statistical Methods for Rates and Proportions*, (3rd ed.). New York: Wiley Series in Probability and Statistics.

Folsom, R.E., and Singh, A.C. (2000). The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification. *Proceedings of the Section on Survey Research Methods* (pp. 598–603). The American Statistical Association.

Haggerty, C., Dugoni, B., Reed, L., Cederlund A, and Taylor, J.R. (1996). *NELS:88/94 Methodology Report* (NCES 96-174). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Hambleton, R.K., Swaminathan, H., and Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, Inc.

Hurst, D., Tan, A., Meek, A., and Sellers, J. (2003). *Overview and Inventory of State Education Reforms: 1990 to 2000* (NCES 2003-020). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J. (1996). *Sample Exclusion in NELS:88—Characteristics of Base Year Ineligible Students: Changes in Eligibility Status After Four Years* (NCES 96-723). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Burns, L.J., Chen, X., Cataldi, E.F., and Charleston, S. (2005a). *A Profile of the American High School Sophomore in 2002* (NCES 2005-338). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Curtin, T.R., Kaufman, P., Alt, M.N., and Chen, X. (2002). *Coming of Age in the 1990s: The Eighth-Grade Class of 1988 12 Years Later* (NCES 2002-321). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Dowd, K.L., Baldridge, J.D., Stipe, J.L., Bartot, V.H., and Frankel, M.R. (1994a). *NELS:88 Second Follow-up: Student Component Data File User's Manual* (NCES 94-374). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Ingels, S.J., Dowd, K.L., Taylor, J.R., Bartot, V.H., Frankel, M.R., and Pulliam, P.A. (1995). *NELS:88 Second Follow-Up: Transcript Component Data File User's Manual* (NCES 95-377). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Ingels, S.J., Planty, M., and Bozick, R. (2005). *A Profile of the American High School Senior in 2004: A First Look—Initial Results From the First Follow-up of the Education Longitudinal Study of 2002 (ELS:2002)* (NCES 2006-348). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ingels, S.J., Pratt, D.J., Rogers, J., Siegel, P.H., and Stutts, E.S. (2004). *Education Longitudinal Study of 2002: Base Year Data File User's Manual* (NCES 2004-405). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., and Stutts, E.S. (2005b). *Education Longitudinal Study of 2002: Base-Year to First Follow-up Data File Documentation* (NCES 2006-344). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M., and Rasinski, K. (1994b). *NELS:88 First Follow-up Final Technical Report* (NCES 94-632). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Kaufman, P., and Rasinski, K. (1991). *Quality of the Responses of Eighth-Grade Students in NELS:88* (NCES 91-487). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Kish, L., and Frankel, M.R. (1974). Inference from Complex Samples. *Journal of the Royal Statistical Society*, Series B (Methodological), *36*: 2–37. Reprinted: G. Kalton and S. Heeringa, eds., *Leslie Kish: Selected Papers* (New York: Wiley, 2003).

Kolen, M.J., and Brennan, R.L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*, (2nd ed.). New York: Springer-Verlag.

Lemke, M., Lippman, L., Bairu, G., Calsyn, C., Kruger, T., Jocelyn, L., Kastberg, D., Liu, Y., Roey, S., and Williams, T. (2001). *Outcomes of Learning: Results from the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy* (NCES 2002-115). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., and Jocelyn, L. (2004). *International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results from the U.S. Perspective*. (NCES 2005-003). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Linn, R.L. (1993). Linking Results of Distinct Assessments. *Applied Measurement in Education*, 6: 83–102.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.

McLaughlin, D.H., and Cohen, J. (1997). *NELS:88 Survey Item Evaluation Report* (NCES 97-052). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Mislevy, R.J. (1992). *Linking Educational Assessments: Concepts, Issues, Methods and Prospects*. Princeton, NJ: Policy Information Center, Educational Testing Service.

Muraki, E.J., and Bock, R.D. (1991). PARSCALE: Parameter Scaling of Rating Data (computer software). Chicago, IL: Scientific Software, Inc.

National Assessment Governing Board (NAGB) (2004). *Mathematics Framework for the 2005 National Assessment of Educational Progress*. Washington, DC: Author.

Planty, M., Bozick R., and Ingels, S.J. (2006). *Academic Pathways, Preparation, and Performance—A Descriptive Overview of the Transcripts from the High School Graduating Class of 2004* (NCES 2007-316). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Planty, M., and DeVoe, J.F. (2005). *An Examination of the Conditions of School Facilities Attended by 10th-Grade Students in 2002* (NCES 2006-302). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Pommerich, M., and Dorans, N.J. (2004). Linking Scores via Concordance. *Applied Psychological Measurement*, 28(4).

Rasinski, K., Ingels, S.J., Rock, D.A., and Pollack, J. (1993). *America's High School Sophomores: A Ten-Year Comparison* (NCES 93-087). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Riccobono, J.A., Henderson, L.B., Burkheimer, G.J., Place, C., and Levinsohn, J.R. (1981). *National Longitudinal Study: Base Year (1972) through Fourth Follow-up (1979) Data File User's Manual.* U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Robitaille, D.F., Schmidt, W.H., Raizen, S., McKnight, C., Britton, E., and Nicole, C. (1993). *Curriculum Frameworks for Mathematics and Science* (TIMSS Monograph No. 1). Vancouver, British Columbia: Pacific Educational Press.

Rock, D.A. (2007). *A Note on Gain Scores and Their Interpretation in Developmental Models Designed to Measure Change in the Early School Years*. (Report RR-07-08). Princeton, N.J.: Educational Testing Service. Available: http://www.ets.org/research/contact.html

Rock, D.A., and Pollack, J.M. (1995). *Psychometric Report for the NELS:88 Base Year Through Second Follow-Up* (NCES 95-382). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Samejima, F. (1994). Roles of Fisher-type information in latent trait models. In H. Bozdogan (Ed.), *Multivariate Statistical Modeling* (Volume 2 of the Proceedings of the First U.S.-Japan Conference on the Frontiers of Statistical Modeling). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Scott, L.A. (2004). *School Library Media Centers: Selected Results From the Education Longitudinal Study of 2002 (ELS:2002)* (NCES 2005-302). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Scott, L.A., and Ingels, S.J. (2007*). Interpreting 12th-Graders' NAEP-Scaled Mathematics Performance Using High School Predictors and Postsecondary Outcomes from the National Education Longitudinal Study of 1988 (NELS:88)* (NCES 2007-328). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Seastrom, M. (2003). *NCES Statistical Standards* (NCES 2003-601). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Available: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003601.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K., and Tourangeau, R. (1990). *NELS:88 Base Year Sample Design Report* (NCES 90-463). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Spencer, B.D., Sebring, P., and Campbell, B. (1987). *High School and Beyond Third Follow-up (1986) Sample Design Report* (NCES 88-402). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

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.

Van der Linden, W.J., and Hambleton, R.K. (Eds.) (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.

Wolter, K. (2007). *Introduction to Variance Estimation*. Second Edition. New York: Springer-Verlag.

Woodruff, R.S. (1971). A Simple Method for Approximating the Variance of a Complicated Estimate. *Journal of the American Statistical Association*, *66*: 411–414.

Zahs, D., Pedlow, S., Morrissey, M., Marnell, P., and Nichols, B. (1995). *High School and Beyond Fourth Follow-up Methodology Report* (NCES 95-426). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.