NCES
National Center for Education Statistics

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

## User's Manual

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

This page is intentionally blank.

NCES
National Center for Education Statistics

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

## User's Manual

**February 2004**

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

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

Jeffrey A. Owings
*Project Officer*
**National Center for Education Statistics**

**U.S. Department of Education**
Rod Paige
*Secretary*

**Institute of Education Sciences**
Grover J. Whitehurst
*Director*

**National Center for Education Statistics**
Robert Lerner
*Commissioner*

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

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

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

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

February 2004

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

**Suggested Citation**

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

**For ordering information on this report, write:**

> U.S. Department of Education
> ED Pubs
> P.O. Box 1398
> Jessup, MD 20794–1398

Or call toll free 1–877–4ED–Pubs

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

# Foreword

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

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

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

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

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

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

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

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

This page is intentionally blank.

# Acknowledgments

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

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

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

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

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

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

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

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

# Table of Contents

# List of Tables

# List of Figures

This page is intentionally blank.

# Chapter 1
# Introduction

## 1.1 Overview of the Data File User's Manual

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

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

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

## 1.2 Historical Background

### 1.2.1 NCES Education High School Longitudinal Studies Program

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

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

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

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

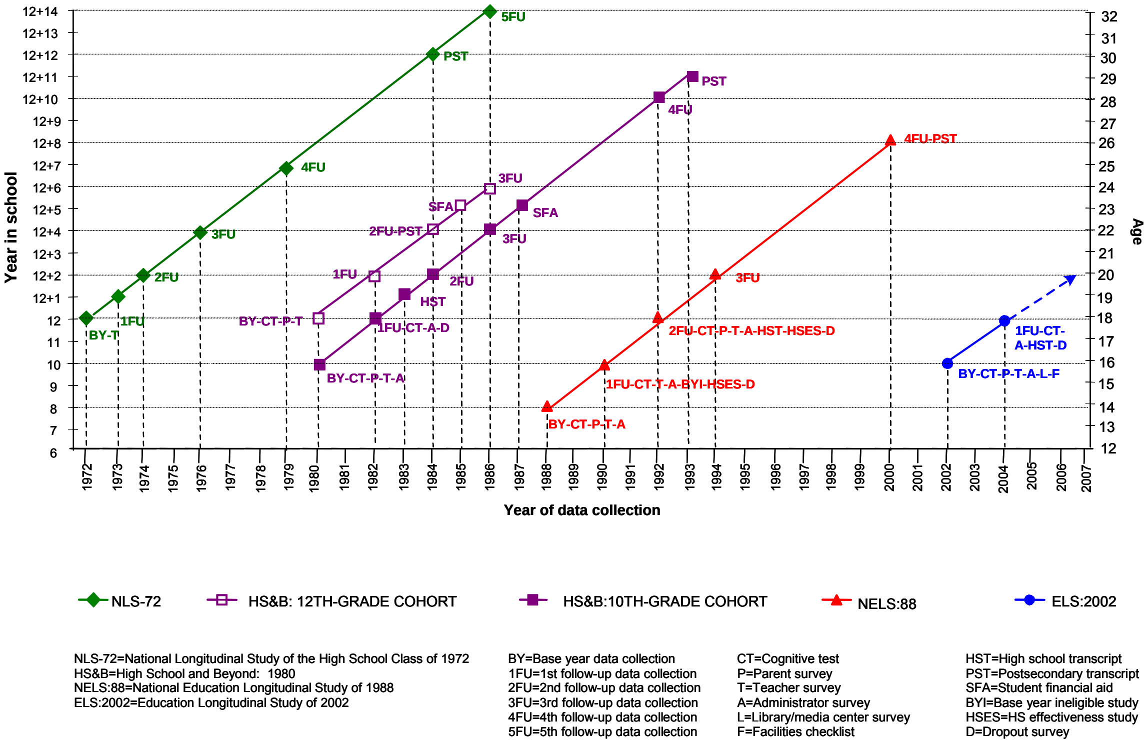
The high school longitudinal studies program consists of three completed studies: the National Longitudinal Study of the High School Class of 1972 (NLS-72), the High School and Beyond (HS&B) longitudinal study of 1980, and the National Education Longitudinal Study of 1988 (NELS:88). In addition, base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

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

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

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

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




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

### 1.2.3 High School and Beyond (HS&B)

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

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

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

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

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

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

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

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

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

#### 1.2.4.1 Base Year

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

#### 1.2.4.2 First Follow-up

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

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

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

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

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

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

#### 1.2.4.3 NELS:88 Second Follow-up

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

#### 1.2.4.4 NELS:88 Third Follow-up

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

#### 1.2.4.5 NELS:88 Fourth Follow-up

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

## 1.3 Education Longitudinal Study of 2002 (ELS:2002)

The base year of ELS:2002 represents the first stage of a major longitudinal effort designed to provide trend data about critical transitions experienced by students as they proceed through high school and into postsecondary education or their careers. The 2002 sophomore cohort will be followed, initially at 2-year intervals, to collect policy-relevant data about educational processes and outcomes, especially as such data pertain to student learning, predictors of dropping out, and high school effects on students' access to, and success in, postsecondary education and the work force.

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

### 1.3.1 ELS:2002 Study Objectives

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

ELS:2002 has two distinctive features. First, it is a longitudinal study, in which the same units are surveyed repeatedly over time. Individual students will be followed for more than 10 years; the base year schools will be surveyed twice, in 2002 (completed) and in 2004. Second, in the high school years, it is an integrated multilevel study that involves multiple respondent populations. The respondents include students, their parents, their teachers, and their schools (from which data have been collected at three levels: from the principal, the librarian, and from a facilities checklist). Each of these two features—the longitudinal nature of the ELS:2002 design and its multilevel focus—will be explained in greater detail below.

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

ELS:2002 gathers information at multiple levels. It obtains information not only from students and their school records, but also from students' parents, teachers, and the administrators (principal and library media center director) of their schools. Data from their teachers, for example, provide information both about the student and the teacher's backgrounds and activities. This multilevel focus supplies researchers with a comprehensive picture of the home, community, and school environments and their influences on the student. This multiple respondent perspective is unified by the fact that, for most purposes, the student is the basic unit of analysis.[7]

Using this multilevel and longitudinal information, the base year (2002) and first follow-up (2004) of ELS:2002 will help researchers and policymakers explore and better understand such issues as the importance of home background and parental aspirations for a child's success; the influence of different curriculum paths and special programs; the effectiveness of different high schools; and whether a school's effectiveness varies with its size, organization, climate or ethos, curriculum, academic press, or other characteristics. These data will facilitate understanding of the impact of various instructional methods and curriculum content and exposure in bringing about educational growth and achievement.

After the high school years, ELS:2002 will continue to follow its sample of students into postsecondary education and/or the labor market. For students who continue on to higher education, researchers can use ELS:2002 to measure the effects of their high school careers on subsequent access to postsecondary institutions, their choices of institutions and programs, and as time goes on, their postsecondary persistence, attainment, and eventual entry into the labor force and adult roles. For students who go directly into the work force (whether as dropouts or high school graduates), ELS:2002 will be able to determine how well high schools have prepared these students for the labor market and how they fare within it.

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

*Base Year (2002)*

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

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

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

*First Follow-up (2004)*

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

*Second Follow-up (2006)*

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

*Further Follow-ups*

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

### 1.3.2 ELS:2002 Research and Policy Issues

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

- students' academic growth in mathematics;
- the process of dropping out of high school—determinants and consequences;
- the role of family background and the home education support system in fostering students' educational success;
- the features of effective schools;
- the impact of coursetaking choices on success in the high school years (and thereafter);
- the equitable distribution of educational opportunities as registered in the distinctive school experiences and performance of students from various subgroups. Such subgroups include:

  - students in public and private high schools;
  - language minority students;
  - students with disabilities;
  - students in urban, suburban, and rural settings;
  - students in different regions of the country;
  - students from upper, middle, and lower socioeconomic status levels;
  - male and female high school students; and
  - students from different racial or ethnic groups.
- steps taken to facilitate the transition from high school to postsecondary education or the world of work.

After ELS:2002 students have completed high school, a new set of issues can be examined. These issues include:

- the later educational and labor market activities of high school dropouts;
- the transition of those who do not go directly on to postsecondary education or to the world of work;
- access to and choice of, undergraduate and graduate educational institutions;
- persistence in attaining postsecondary educational goals;
- rate of progress through the postsecondary curriculum;
- degree attainment;
- barriers to persistence and attainment;
- entry of new postsecondary graduates into the work force;
- social and economic rate of return on education to both the individual and society; and
- adult roles, such as family formation and civic participation.

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

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

#### 1.3.2.1 Cross-sectional Profiles

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

#### 1.3.2.2 Longitudinal Analysis

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

#### 1.3.2.3 Intercohort Comparisons

As part of an important historical series of studies that repeats a core of key items each decade, ELS:2002 offers the opportunity for the analysis of trends in areas of fundamental importance, such as patterns of coursetaking, rates of participation in extracurricular activities, academic performance, and changes in goals and aspirations. A 1980–2002 NCES high school sophomore trend report is currently in preparation. With completion of the first follow-up in 2004, researchers will be able to compare ELS:2002 high school seniors' experience, attitudes, and achievement with that of NELS:88 seniors in 1992, HS&B seniors in 1980, and NLS-72 seniors in 1972. They will also be able to compare ELS:2002 dropouts in 1984 with the high school dropouts studied by HS&B in 1982 and by NELS:88 in 1992. Such cross-cohort comparisons are of particular importance to measuring the nation's goals in achieving equity in educational opportunities and outcomes and in measuring the success of school reform and related initiatives.

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

#### 1.3.2.4 International Comparisons

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

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

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

### 1.3.3 Overview of the Base Year Study Design

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

Seven study components comprise the base year design: assessments of students (achievement tests in mathematics and reading); a survey of students; surveys of parents, teachers, school administrators, and librarians; and a facilities checklist (completed by survey administrators, based on their observations at the school). The student assessments measured achievement in mathematics and reading; the baseline scores can serve as a covariate or control variable for later analyses. Mathematics achievement will be reassessed 2 years hence, so that achievement gain over the last 2 years of high school can be measured and related to school processes and mathematics coursetaking. The student questionnaire gathered information about the student's background, school experiences and activities, plans and goals for the future, employment and out-of-school experiences, language background, and psychological orientation toward learning.

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

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

# Chapter 2
# Instrumentation

## 2.1 Introduction

The data collection instruments for the Education Longitudinal Study of 2002 (ELS:2002) base year consisted of five separate questionnaires (student, parent, teacher, school administrator, and library media center), two achievement tests (assessments in reading and mathematics), and a school observation form (facilities checklist). The base year questionnaires can be found in the electronic version of this data file user's manual (appendix B) as PDF files on the NCES ELS:2002 web site (*http://nces.ed.gov/surveys/els2002/*).

### 2.1.1 Instrument Development Process and Procedures

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

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

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

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

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

### 2.1.2 Instrument Development Goals and Constraints

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

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

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

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

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

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

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

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

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

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

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

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

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

## 2.2 Student Questionnaire and Assessments

### 2.2.1 Student Questionnaire

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

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

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

By far the longest section of the student questionnaire was the module on school experiences and activities. The principal content strands in this section inquire about school climate, student recognition, school disengagement behaviors (tardiness, classes-cutting, etc.), perception of high school program placement (academic, general, or vocational track), attitudes toward school and motivation for attending school, learning environment of the math class, use of computer technology, receipt of special services, time spent on homework, importance of grades to the student, school-sponsored activities (sports and extracurricular activities), time spent in reading and outside activities (including television viewing and video games), and use of the library media center. There are also questions (with parallels on the parent instrument) about close school friends and their parents that are intended, among other uses, to measure aspects of embeddedness in social networks that might be a source of social capital.

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

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

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

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

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

### 2.2.2 Reading and Mathematics Assessments

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

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

#### 2.2.2.1 Test Design and Format

Test specifications for ELS:2002 were adapted from frameworks used for NELS:88. Math tests contained items in arithmetic, algebra, geometry, data/probability, and advanced topics and were divided into process categories of skill/knowledge, understanding/ comprehension, and problem solving. The ELS:2002 math tests placed a somewhat greater emphasis on practical applications and problem solving than did the NELS:88 test forms. Reading tests consisted of reading passages of one paragraph to one page in length, followed by three to six questions based on each passage. The reading passages included literary material as well as topics in the natural and social sciences. Several passages required interpretation of graphs. Questions were categorized as reproduction of detail, comprehension, or inference/evaluation.

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

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

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

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

#### 2.2.2.2 Scoring Procedures

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

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

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

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

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

#### 2.2.2.3 Score Descriptions and Summary Statistics

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

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

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

The database also reports scores for ELS:2002 participants on the mathematics score scale used for NELS:88 10th graders in 1990[13] and on the PISA:2000 reading scale.

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

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

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

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

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

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

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

**Standardized scores (T-scores)**. T-scores provide norm-referenced measurements of achievement; that is, estimates of achievement level relative to the population as a whole. A high mean T-score for a particular subgroup indicates that the group's performance is high in comparison to other groups. It does not represent mastery of a particular set of skills, only that the subgroup's mastery level is greater than a comparison group. In other words, T-scores provide information on status compared to students' peers, while the IRT-estimated number-right scores represent status with respect to achievement on a particular criterion set of test items. The T-scores can only provide an indicator of the extent to which an individual or a subgroup ranks higher or lower than the national average.

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

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

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

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

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

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

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

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

#### 2.2.2.4 Links to Other Surveys

Scores for ELS:2002 are reported on scales that permit comparisons with reading and mathematics data for NELS:88 10th graders, and with PISA:2000 reading results for 15-year-olds. The link to the NELS:88 scales represents a "true" equating. This means that the tests may be considered interchangeable or, in other words, a score on one exam should be equivalent to a score on the other exam. Several conditions must be met for equating two tests. Most importantly, the tests must measure the same content. Similarity of format, length, reliability, and subgroup performance also support the interpretation of interchangeable scores. The differences between ELS:2002 and PISA, described below, were sufficiently important that the PISA scale scores represent a concordance, or a link based on population distributions, rather than equivalent or interchangeable scores.

**NELS:88-equated Scores.** Equating the ELS:2002 scale scores to the NELS:88 scale scores was completed through common-item or *anchor equating*. The ELS:2002 and NELS:88 tests shared 30 reading and 49 math items. These common items provided the link that made it possible to obtain ELS:2002 student ability estimates on the NELS:88 ability scale. (The ELS:2002 data for seven of the reading items and six of the math items did not fit the NELS:88 IRT parameters, so these items were not treated as common items for the purpose of equating.) Parameters for the common items were fixed at their NELS:88 values, resulting in parameter estimates for the noncommon items that were consistent with the NELS scale.

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

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

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

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

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

- *Reading levels:*

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

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

- *Mathematics levels:*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## 2.3 Parent Questionnaire

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

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

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

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

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

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

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

## 2.4 Teacher Questionnaire

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

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

### 2.4.1 Teacher Evaluations of ELS:2002 Students

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

### 2.4.2 Teacher Background

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

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

## 2.5 School Administrator Questionnaire

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

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

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

### 2.5.1 School Characteristics

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

### 2.5.2 Student Characteristics

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

### 2.5.3 Teaching Staff Characteristics

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

### 2.5.4 School Policies and Programs

The module on *school policies and programs* addresses two areas of school policy: competency tests for high school graduation and school safety and crime. Schools that administer competency tests are asked to indicate the grades in which the tests are given; whether the exams are a requirement of the state, district, or the school; the subject areas covered on the tests; and the percentage who fail the exams and the recourse for these students. School administrators are also questioned about security measures taken by the school and efforts to involve parents in school discipline and safety.

### 2.5.5 Technology

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

### 2.5.6 School Governance and Climate

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

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

## 2.6 Library Media Center Questionnaire

A library media center questionnaire was not a feature of ELS:2002's predecessor studies. The addition of this component will permit investigation of the role of integration between library media resources and curriculum and classroom instruction in promoting effective learning. For the school library media center component, the school librarian, media center director, or school administrator supplied information about library media center size, organization, and staffing; technology resources and electronic services; extent of library and media holdings, including both collections and expenditures; and levels of facility utilization, including scheduling for use by students and teachers. Finally, the questionnaire also supplies information about the library media center's use in supporting the school's curriculum, that is, how library media center staff collaborate with and support teachers to help them plan and deliver instruction. Information in the library media center questionnaire can be used as contextual data with the student as the unit of analysis. Or, data from the questionnaire can be used at the school level to generalize to libraries within all regular high schools with 10th grades in the United States in the 2001–02 school year. The ELS:2002 library media center questionnaire is largely an abridgment of the school library media center questionnaire used in SASS:2000.

## 2.7 School Facilities Checklist

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

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

## 3.1 Introduction

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

### 3.1.1 Sample Design

The sample design for the Education Longitudinal Study of 2002 (ELS:2002) is similar in many respects to the designs used in the three prior studies of the National Education Longitudinal Studies Program, the National Longitudinal Study of the High School Class of 1972 (NLS-72), the High School and Beyond (HS&B) longitudinal study, and the National Education Longitudinal Study of 1988 (NELS:88). ELS:2002 is different from NELS:88 in that the ELS:2002 base year sample students are 10th-graders rather than 8th graders. As in NELS:88, there were oversamples of Hispanics and Asians in ELS:2002. However, for ELS:2002, counts of Hispanics and Asians were obtained from the Common Core of Data (CCD) and the Private School Survey (PSS) to set the initial oversampling rates.

ELS:2002 used a two-stage sample selection process. First, schools were selected with probability proportional to size (PPS), and school contacting resulted in 1,221 eligible public, Catholic, and other private schools from a population of approximately 27,000 schools containing 10th-grade students. Of the eligible schools, 752 participated in the study. A full discussion of the sample design and response rates is presented in chapters 3 and 4. These schools were then asked to provide 10th-grade enrollment lists. In the second stage of sample selection, approximately 26 students per school were selected from these lists.

The ELS:2002 base year comprises two primary target populations—schools with 10th grades and 10th-grade students—in the spring term of the 2001–02 school year. Schools and students are intended as the study's basic units of analysis. School-level data reflect a school administrator questionnaire, a library media center questionnaire, a facilities checklist, and the aggregation of student data to the school level. Student-level data consist of student questionnaire and assessment data, and reports from students' teachers and parents. (School-level data, however, can also be reported at the student level and serve as contextual data for students.)

### 3.1.2 Schools

The target population of schools for the full-scale ELS:2002 study consisted of regular public schools, including State Education Agency schools and charter schools, and Catholic and other private schools that contain 10th grades and are in the United States (the 50 states and the District of Columbia).

The sampling frame of schools was constructed with the intent to match the target population. However, selected schools were determined to be ineligible if they did not meet the definition of the target population. Responding schools were those schools that had a Survey Day (i.e., data collection occurred for students in the school).[15] Of the 1,268 sampled schools, 1,221 were eligible. Some 752 of the 1,221 eligible schools responded for a 67.8 percent (weighted[16]) response rate.

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

### 3.1.3 Students

The target population of students for the full-scale ELS:2002 consisted of spring-term 10th graders in 2002 (excluding foreign exchange students) enrolled in schools in the school target population.

The sampling frames of students within schools were constructed with the intent to match the target population. However, selected students were determined to be ineligible if they did not meet the definition of the target population.

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

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

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

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

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

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

### 3.1.4 Parents

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

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

### 3.1.5 Teachers

Teacher data also have been collected to support analyses at the student level. All full- and part-time teachers who were teaching 10th graders in mathematics or English/language arts in spring 2002 were included in the ELS:2002 universe of 10th-grade teachers. The actual sample was restricted to teachers who were providing instruction in one of the two subjects to the full (expanded) sample of 10th-grade students within the sampled schools. Thus, as for parents, there was no need to construct a formal universe list of 10th-grade mathematics and English teachers prior to their selection. Once students were selected within schools, their teachers of the assigned subject pairs were identified and asked to participate in the study. It is important to remember that the student is the central unit of analysis and that teacher reports were collected to provide contextual data for students.

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

### 3.1.6 Nonresponse Bias Analysis

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

### 3.1.7 Imputation

After the editing process (which included logical imputations), the remaining missing values for 14 analysis variables and two ability estimates (reading and mathematics) were statistically imputed. These variables were chosen because they are key variables used in standard reporting and cross-sectional estimation. Imputation of missing values for the ability estimates provided complete information for the various test scores derived from those estimates, including both normative and criterion-referenced scores. The imputations were performed primarily to reduce the bias of survey estimates caused by missing data. The imputed data also made the data more complete and easier to analyze. Most of the variables were imputed using a weighted hot-deck procedure.[18] The order of imputation addressed problems of multivariate association by using a series of univariate models fitted sequentially such that variables modeled earlier in the hierarchy had a chance to be included in the covariate set for subsequent models. Additionally, multiple imputations were used for a few variables, including test scores. Imputation is discussed in detail in section 3.3.

### 3.1.8 Weighting

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

[18] See Cox (1980).

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

### 3.1.9 Standard Errors and Design Effects

The variance estimation procedure had to take into account the complex sample design, including stratification and clustering. One common procedure for estimating variances of survey statistics is the Taylor series linearization procedure. This procedure takes the first-order Taylor series approximation of the nonlinear statistic, and then substitutes the linear representation into the appropriate variance formula based on the sample design. For stratified multistage surveys, the Taylor series procedure requires analysis strata and analysis primary sampling units (PSUs). Therefore, analysis strata and analysis PSUs were created. The impact of the departures of the ELS:2002 complex sample design from a simple random sample design on the precision of sample estimates can be measured by the design effect. Appendix K presents standard errors and design effects for 30 means and proportions based on the ELS:2002 student, parent, and school data for the sample (as a whole and for selected subgroups).

### 3.1.10 Disclosure Analysis

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

## 3.2 Base Year Sample Design

### 3.2.1 Sampling Frame

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

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

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

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

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

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

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

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

To construct a supplemental sampling frame of new schools not currently on the sampling frame, a subsample of the public schools was selected. Each district associated with this new subsample of schools was considered in sample and was asked to identify public schools in their jurisdiction that taught 10th graders, that had recently opened, or that had begun teaching 10th graders in the last 2 years. The districts were provided with a list of all public schools on the sampling frame in their district to help them identify the appropriate schools.

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

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

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

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

### 3.2.2 Stratification and School Selection

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

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

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

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

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

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

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

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

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

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

A sample size larger than 800 schools was necessary to compensate for the anticipated nonresponse.  The sample was randomly divided by stratum into two release pools and a reserve pool.  The two release pools were the basic sample, with the schools in the second pool being released randomly within stratum in waves as needed to achieve the sample size goal.  Also, the reserve pool was released selectively in waves by simple random sampling within stratum for strata with low yield and/or response rates, when necessary.  To determine the overall sample size, assumptions were made about the expected response rate.  Based on historical response rates in NELS:88 and HS&B, an overall response rate of approximately 70 percent was expected, but a response rate greater than 70 percent was attempted.  It was planned to increase the sample size to 1,600, since there was uncertainty about achieving a response rate of at least 70 percent (sample size of 1,143).  Such a response rate turned out to be harder to achieve than was the case in NELS:88 and HS&B (see table 43 in chapter 4).  Table 11 gives the counts and percentages of sampled, eligible, and participating schools.  These results show the difficulty of getting different types of schools to participate.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## 3.2.3 Selection of Students

### 3.2.3.1 Sample Sizes

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

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

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

[21] Sometimes called Poisson sampling.

**Table 13. Domain sample size requirements: 2002**

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

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

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

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

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

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

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

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

### 3.2.3.2 Specifications for Enrollment Lists

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

The information requested for each eligible student was as follows:

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

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

It was requested that the electronic file be a column-formatted or comma-delimited American Standard Code for Information Interchange (ASCII) file or Excel file. Schools provided the electronic lists via e-mail, provided a diskette containing the file, or uploaded the file to the ELS:2002 web site. If the school could not provide an electronic file, it was asked to provide a hardcopy list, preferably in alphabetical order within race/ethnicity strata to facilitate stratified systematic sampling. Whatever the school could provide was accepted to select the student samples; however, every effort was made to facilitate receiving uniformly formatted electronic files from as many schools as possible to make processing them quicker, more reliable, and less expensive.

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

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

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

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

#### 3.2.3.3 Quality Assurance Checks

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

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

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

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

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

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

### 3.2.3.4 Student Sampling from Lists

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

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

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

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

For schools that provided hardcopy lists, an efficient two-stage process was used to select systematic samples from hardcopy lists. Sample pages were first selected and then sample students within strata were chosen from the selected pages. The page sampling rate was set so that approximately 10 students were selected from each page. This approach was particularly efficient for long lists. The sample was keyed after it was selected.

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

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

#### 3.2.3.5 Sample Updating

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

### 3.2.4 Student Eligibility and Exclusion

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

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

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

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

#### 3.2.4.1 Schools Given Clear Criteria for Including/Excluding Students

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

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

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

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

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

#### 3.2.4.2 Accommodations Offered to Increase Participation

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

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

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

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

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

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

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

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

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

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

[25] See Ingels (1996).

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

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

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

### 3.2.5 Selection of Contextual Samples

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

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

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

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

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

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

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

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

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

### 3.2.6 Bias Analysis for School and Student Unit Nonresponse

Unit nonresponse causes bias in survey estimates when the outcomes of respondents and nonrespondents are different. ELS:2002, has two levels of unit response: school response, defined as the school participating in the study by having a Survey Day, on which the students took the test and completed the questionnaires, and student response, defined as the student completing at least a specified portion of the student questionnaire. The final overall school weighted response rate was 67.8 percent, and the final pool 1 weighted response rate was 71.1 percent. The final student weighted response rate was 87.3 percent. Because the school response rate was less than 70 percent in some domains and overall, analyses were conducted to determine if school estimates were significantly biased due to nonresponse. For students, although the overall unweighted response rate was approximately 87 percent, the response rate was below 85 percent for certain domains, so a student level nonresponse bias analysis conditional on the school responding was also conducted. See section 4.9 for a further discussion of response rates.

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

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

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

Some information on characteristics of nonresponding students was available from student enrollment lists. On these lists, data were obtained on IEP status, race/ethnicity, and sex. These data were not provided by all schools (in particular, information on IEP status was often missing, and IEP information was typically relevant only for public schools). In consequence, only the school-supplied race/ethnicity and sex data, as well as the school-level data used in the school nonresponse bias analysis, were utilized in conducting the student-level nonresponse bias analysis.

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

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

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

The nonresponse bias was estimated for variables known for both respondents and nonrespondents. The bias in an estimated mean based on respondents, $\overline{y}_R$, 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 is available. The estimation of π is as follows:

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

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

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

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 SUDAAN software package.

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

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

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

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

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

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

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

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

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

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

See notes at end of table.

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

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

See notes at end of table.

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

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

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

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

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

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

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

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

See notes at end of table.

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

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

See notes at end of table.

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

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

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

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

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

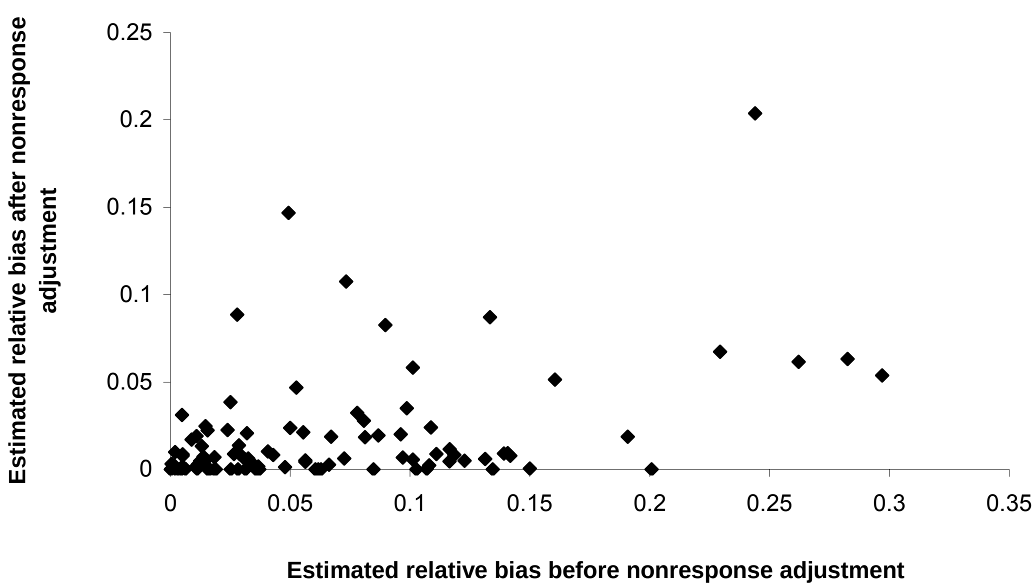
* Statistically significant at the 0.05/(c-1) level, where c is the number of categories within the primary variable.
[1] Design weight is used before nonresponse adjustment. This is the mean of the continuous variable.
[2] Weight after nonresponse adjustment.
[3] IEP = Individualized education program.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

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

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

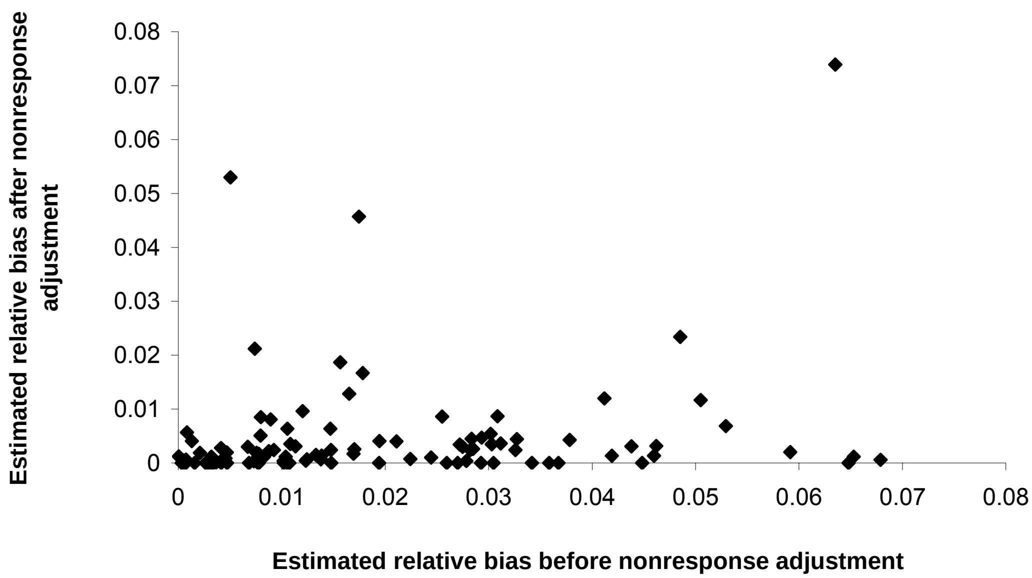
No additional nonresponse bias analysis was necessary to account for nonresponse from school administrators, libraries, or facility checklists because each of these had a response rate greater than 95 percent.

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



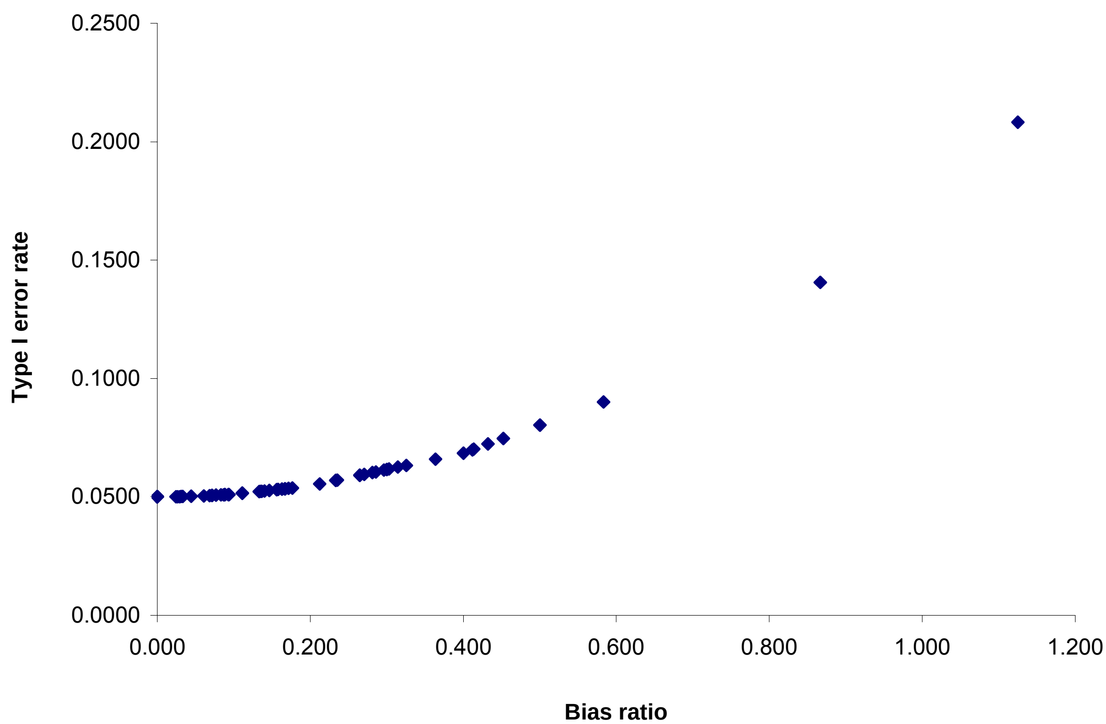

SOURCE: U.S. Department of Education, National Center for Education Statistics, Education Longitudinal Study of 2002 (ELS:2002).

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




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

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




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

In conclusion, examination of variables known for most respondents and nonrespondents before nonresponse adjustment detected some degree of bias. The school and student nonresponse bias analyses in conjunction with the weighting adjustments described in section 3.4 were not successful in eliminating all bias. However, they reduced bias and eliminated significant bias for the variables known for most respondents and nonrespondents, which were considered to be some of the more important classification and analysis variables. The data user should exercise caution in using the data because bias was not estimated and corrected for all variables. The relative bias decreased considerably after weight adjustments—especially when it was large before nonresponse adjustment, and the relative bias remained small after weight adjustments when it was small before nonresponse adjustment usually.

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

### 3.2.7 Questionnaire Item Nonresponse Bias Analysis

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

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

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

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

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

#### 3.2.7.1 High Nonresponse Questionnaire Variables: Student-Level Items

No parent or teacher questionnaire items fell below 85 percent response. However, there were 78 such items on the student questionnaire, including composites. Item nonresponse was an issue for the student questionnaire because, in timed sessions, not all students reached the final items. Student-survey item nonresponse is primarily a function of questionnaire position, with the highest nonresponse seen in the final item, which was answered by only 64.6 percent of respondents. The 78 student variables evidencing high (>15 percent) nonresponse[29] are listed in table 24.

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

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

See note at end of table.

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

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

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

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

### 3.2.7.2 High Nonresponse Questionnaire Variables: School-Level Items

At the school level, 41 administrator items fell below 85 percent (ranging from 84.7 percent to a low of 74.6 percent). No library media center questionnaire items fell below the threshold, nor did any facility checklist items fall below 85 percent. While the school-level items will often be used as contextual data with the student as the basic unit of analysis, these items are also, with the school weight, generalizable at the school level. Therefore, for the school administrator questionnaire, nonresponse rates and nonresponse bias estimates have been produced at the school level. While item nonresponse in the student questionnaire reflects item position in the questionnaire and the inability of some students to reach the final items in a timed session, nonresponse in the school questionnaire must be explained by two other factors. First, the nature of particular items, and second, the fact that some administrators completed an abbreviated version of the school administrator questionnaire (the high nonresponse items did not appear on the abbreviated instrument).

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

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

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

See note at end of table.

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

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

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

### 3.2.7.3 Estimating Bias

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

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

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

Despite the limitations imposed by some missingness, these variables are hypothesized to be especially helpful in explaining student questionnaire nonresponse and its biases. Since, on the student questionnaire, nonresponse is primarily a function of question position, one may hypothesize that poor readers in particular (or poor students more generally) are most likely to be missing on the final items. An additional reason for including both the math and reading quartile is that a number of the high nonresponse student variables have to do with psychological orientations toward mathematics or English, such that, for these scales, any bias by reading or mathematics achievement level would be particularly interesting to quantify.

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

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

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

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

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

| | |
|---|---|
| Overall mean | 1.20 |
| Median | 0.75 |
| Standard deviation | 1.19 |

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

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

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

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

[30] Appendix I can be found in the electronic version of this user's manual on the NCES web site, in the form of a PDF file (*http://nces.ed.gov/surveys/els2002/*).
[31] In addition to 35 characteristics, for four imputed variables, a holding category of "imputed" is also supplied (sex, race, mother's educational attainment, and home language). In a fifth instance, two Hispanic categories (race specified/not specified) are combined for one overall Hispanic ethnicity category.

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

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

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

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

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

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

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

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

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

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

| | |
|---|---|
| Overall mean | 1.12 |
| Median | 0.84 |
| Standard deviation | 1.11 |

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Apart from the caveats that must be entered about the comparability of the ELS:2002 item, it is also important to consider that the ELS:2002 format may have been open to misinterpretation by some respondents. (This observation is speculative; it is not based on cognitive interviews with 10th graders, or re-interviews of ELS:2002 respondents.) In particular, although the question stems say, “how many hours a day,” splitting the response boxes into weekdays and weekends may have abetted some respondents in the error of reporting total weekday and total weekend hours. If some students forgot the definition in the question stem ("how many hours per day") and misinterpreted "weekdays" as the total number of hours on weekdays in a week, an inflated estimate for high-end use would be the likely consequence. A parallel error could be made for the "on weekends" portion of the question. Estimates from television viewing items in the past have been quite sensitive to small format differences (see Rasinski, Ingels, Rock, and Pollack 1993, appendix B, pp. 15–18). While reliable comparison sources are not available for the video game item, one may presume that because it was identical in format to the television viewing item, it would be open to a like degree of respondent error, and that that error would be in the same direction (i.e., somewhat inflated high-end estimates).

## 3.3 Imputation

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

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

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

[32]See Cox (1980).

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

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

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

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

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

Weighted sequential hot deck imputation has as an advantage that it controls the number of times that a respondent record can be used for imputation and gives each respondent record a chance to be selected for use as a hot deck donor. To implement the weighted hot deck procedure, imputation classes and sorting variables that were relevant for each item being imputed were defined. If more than one sorting variable was chosen, a serpentine sort was performed where the direction of the sort (ascending or descending) changed each time the value of a variable changed. The serpentine sort minimized the change in the student characteristics every time one of the variables changed its value.

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

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

## 3.4 Calculation of Weights and Results of Weighting

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

### 3.4.1 School Weights

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

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

where:

$n_r$ = the sample size in stratum r,

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

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

Therefore, the school sampling weight was assigned as follows:

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

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

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

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

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

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

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

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

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

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

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

A responding school was defined as a school that had a Survey Day. There were 752 such schools.[36] A list of all variables that were for the generality of both responding and nonresponding schools (these variables, some taken from the sampling frame and others collected from schools and districts, are listed in section 3.2.6) were compiled. For data known for most, but not all, schools that would be useful to include in the nonresponse adjustment, weighted hot-deck imputation was used so that there would be data for all eligible sample schools. Then, these variables were main effects in the model. The weight specified for the

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

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

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

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

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

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

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

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

See notes at end of table.

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

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

See notes at end of table.

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

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

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

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

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

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

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

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

See notes at end of table.

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

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

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

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

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

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

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

### 3.4.2 Student Weights

Two sets of student weights were computed. There is one set of weights for student questionnaire completion (BYWTSTU); this is the sole student weight that appears on the public-use file and generalizes to the population of spring 2002 sophomores who were capable of completing an ELS:2002 student questionnaire. A second set of weights for the expanded sample of questionnaire-eligible and questionnaire-ineligible students appears only on the restricted-use file. This weight sums to the total of all 10th-grade students.[37]

First, the student-level design weight (WT6) was calculated. The sample students were systematically selected from the enrollment lists at school-specific rates that were inversely proportional to the school's probability of selection. Specifically, the sampling rate for student stratum s within school I ($f_{s\backslash i}$)was calculated as the overall sampling rate divided by the school's probability of selection, or:

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

where:

$f_s$ = the overall student sampling rate, and

$\pi_r$ (i) = the school's probability of selection.

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

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

selected did not exceed 35 students. A minimum sample size constraint of 10 students also was imposed, if a school had more than 10 tenth graders. Adjustments to the sampling rates were also made (as sampling progressed) to increase the sample size in certain student strata that were falling short of the sample size targets.

The student sampling weight then was calculated as the reciprocal of the school-specific student sampling rates, or:

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

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

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

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

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

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

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

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

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

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

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

See notes at end of table.

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

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

See notes at end of table.

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

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

See notes at end of table.

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

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

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

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

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

See notes at end of table.

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

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

See notes at end of table.

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

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

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

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

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

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

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

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

### 3.4.3 Quality Control

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

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

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

## 3.5 Standard Errors and Design Effects

### 3.5.1 Standard Errors

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

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

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

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

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

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

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

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

### 3.5.2 Design Effects

The impact of the departures of the ELS:2002 complex sample design from a simple random sample design on the precision of sample estimates can be measured by the design effect.[44] The design effect is the ratio of the actual variance of the statistic to the variance that would have been obtained had the sample been a simple random sample. The design standard errors will be different from the standard errors that are based on the assumption that the data are from a simple random sample. The ELS:2002 sample departs from the assumption of simple random sampling in three major respects: both schools and student samples were stratified by school and student characteristics, respectively; both schools and students were selected with unequal probabilities of selection; and the sample of students was clustered by school. A simple random sample is, by contrast, unclustered and not stratified. Additionally, in a simple random sample, all members of the population have the same probability of selection. Generally, clustering and unequal probabilities of selection increase the variance of sample estimates relative to a simple random sample, and stratification decreases the variance of estimates.

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

Standard errors and design effects were computed at the first stage (school level) and at the second stage (student level). There are multiple instruments at both levels. At the school level, there was a school administrator questionnaire, a library media center questionnaire, and a facilities checklist. The school administrator questionnaire was the basis for the school-level calculations, however, two items from the library questionnaire were also included. At the school level, there were student, parent, and teacher questionnaires. For student-level calculations, items from both the student and parent questionnaires were used. Therefore, three sets of standard errors and design effects were computed (school, student, and parent), and this is similar to what was done for NELS:88. Each of the three sets includes standard errors and design effects for 30 means and proportions overall and for subgroups.

The school subgroups are similar to those used in NELS:88:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

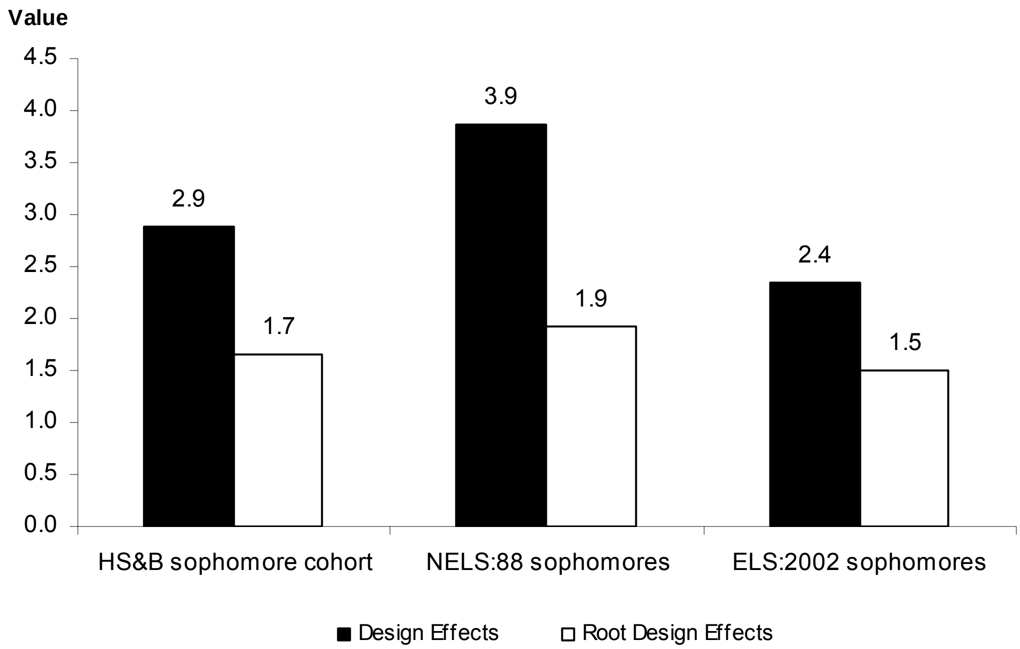
**Table 41. Mean design effects (DEFFs) and root design effects (DEFTs) for parent questionnaire data: 2002**

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

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

The student-level design effects indicate that the ELS:2002 sample was more efficient than the NELS:88 sample and the High School and Beyond (HS&B) sample. For means and proportions based on student questionnaire data for all students, the average design effect in ELS:2002 was 2.35; the comparable figures were 3.86 for NELS:88 sophomores and 2.88 for the HS&B sophomore cohort. This difference is also apparent for some subgroup estimates. Ingels et al. (1994) present design effects for 16 subgroups defined similarly to those in table 40 above. For all 16 subgroups, the ELS:2002 design effects are smaller on the average than those for the NELS:88 sophomores. Frankel et al. (1981) also present design effects for eight subgroups defined similarly to those in table 40 above. For all eight subgroups, the ELS:2002 design effects are smaller on average than those for the HS&B sophomore cohort. Figure 5 shows the mean design effects and root design effects for HS&B sophomores, NELS:88 sophomores, and ELS:2002 sophomores.

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




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

The smaller design effects in ELS:2002 compared to those for NELS:88 sophomores are probably due to disproportional strata representation introduced by subsampling in the NELS:88 first follow-up. Although the general tendency in longitudinal studies is for design effects to lessen over time, as dispersion reduces the original clustering, subsampling increases design effects. This is because subsampling introduces additional variability into the weights with an attendant loss in sample efficiency.

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

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

where $\bar{b}$ refers to the average cluster size (the average number of students selected from each school) and rho refers to the intraclass correlation coefficient, a measure of the degree of within-cluster homogeneity. If the value of rho (which varies from one variable to the next) averaged about 0.05 in both studies, then the reduced cluster size in ELS:2002 would almost exactly account for the reduction in the design effects relative to HS&B.

The ELS:2002 parent-level design effects are similar to the student-level design effects. For estimates applying to all students, the average design effect was 2.24 for the parent data and 2.35 for the student data. For almost all subgroups, the average design effect was lower for the parent data than for the student data.

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

The school-level design effects reflect only the impact of stratification and unequal probabilities of selection because the sample of schools was not clustered. Therefore, it could be expected that the design effects for estimates based on school data would be small compared to those for estimates based on student and parent data. However, this is not the case, as the school average design effect is 2.76. The reason for this is that the sample was designed to estimate students with low design effects. In addition to stratifying schools, a composite measure of size was used for school sample selection based on the number of students enrolled by race (see section 3.2.2). This is different from the methodology used for NELS:88 (see Spencer et al. 1991). The NELS:88 average school design effect (in the base year study) was considerably lower, 1.82.

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

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

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

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

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

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

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

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

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

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

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

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

## 3.6 Disclosure Risk Analysis and Protections

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

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

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

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

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

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