NATIONAL CENTER FOR EDUCATION STATISTICS

User's Manual **July 2002**

National Education Longitudinal Study of 1988

Base-Year to Fourth Follow-up Data File User's Manual

U.S. Department of Education
Office of Educational Research and Improvement **NCES 2002-323**

NATIONAL CENTER FOR EDUCATION STATISTICS

User's Manual **July 2002**

National Education Longitudinal Study of 1988

Base-Year to Fourth Follow-up Data File User's Manual

Thomas R. Curtin
Steven J. Ingels
Shiying Wu
Ruth Heuer
Research Triangle Institute

Jeffrey Owings
Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement **NCES 2002-323**

U.S. Department of Education
Rod Paige
Secretary

Office of Educational Research and Improvement
Grover J. Whitehurst
Assistant Secretary

National Center for Education Statistics
Gary W. Phillips
Deputy 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
Office of Educational Research and Improvement
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

July 2002

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

Suggested Citation
Curtin, T.R., Ingels, S.J., Wu, S., and Heuer, R. (2002). *National Education Longitudinal Study of 1988: Base-Year to Fourth Follow-up Data File User's Manual (NCES* 2002-323). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

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 or go to the Internet: ***http://www.ed.gov/pubs/edpubs.html***

Contact: Jeffrey Owings, Project Officer, (202) 502-7423

Preface

This manual has been produced to familiarize data users with the procedures followed for data collection and processing in the National Education Longitudinal Study of 1988 (NELS:88). A corollary objective has been to provide the necessary documentation for use of the data file.

This manual provides information about NELS:88 from beginning (its 1988 base year) to end (the fourth follow-up data collection in 2000). However, because the base year through third follow-up rounds were documented in a series of prior user's manuals and methodology reports, the authors have attempted to supply somewhat more comprehensive information for the fourth follow-up round. In addition, while some information is provided about restricted use data, the primary focus of this manual is on the public use data, particularly as contained in the public use Electronic Codebooks (ECBs). The manual contains five chapters and six appendices.

Chapter 1 serves as an introduction to NELS:88. It provides an overview of the study, including its historical context. It summarizes the objectives of each round of the study and points to further sources of documentation for the prior rounds.

Chapter 2 supplies information about the NELS:88 data collection instruments. It includes discussions of (1) the various NELS:88 questionnaires; (2) the achievement tests in mathematics, science, reading, and social studies that comprise the NELS:88 cognitive test battery; and (3) such archival records data as were gathered in the high school and postsecondary transcript components of the study.

Chapter 3 addresses the NELS:88 sample design, weighting, and design effects. Base year school and student selection, subsampling and freshening over the in-school years, and further subsampling in the out-of-school (1994 and 2000) rounds are among the topics discussed. This chapter also describes weighting procedures, and documents the design effects for each wave of data.

Chapter 4 provides an account of data collection methodology and results for all NELS:88 waves and respondent populations, while chapter 5 documents data processing procedures.

Finally, there are six appendices. The first appendix (appendix A) provides a concise guide to using the NELS:88 data. Appendix B provides information about data that is available only on a restricted use basis. Appendix C supplies tables of response rates for all rounds of the study. Appendix D provides standard error and design effect tables for the fourth follow-up round. Appendix E is a glossary of NELS:88 terms. Appendix F lists important derived variables from the fourth follow-up data collection in 2000.

Acknowledgments

Many people contributed to the production of this manual, and the authors would like to thank several in particular. We especially appreciate the ongoing and patient guidance provided by Jeff Owings, the NELS:88 fourth follow-up study's Project Officer at the U.S. Department of Education, National Center for Education Statistics (NCES). Thanks, too, to Cliff Adelman, Office of Educational Research and Improvement (OERI), for his efforts in behalf of the NELS:88 postsecondary education transcript study.

We would also like to acknowledge the helpful comments of reviewers of this user's manual. Bruce Daniel (Pinkerton Computer Consultants), Marty Frankel (Abt Associates), Anne Hafner (California State University, Los Angeles), Ken Rasinski (National Opinion Research Center), and Leslie Scott (Education Statistics Services Institute) all provided useful comments and suggestions, as did several NCES reviewers: Ilona Berkovits, Dennis Carroll, Chris Chapman, Ralph Lee, and Sam Peng. Bruce Taylor served as the adjudicator. Our thanks to these individuals for the quality of their reviews and the helpful guidance that they provided.

We also wish to thank the members of the NELS:88 fourth follow-up study's technical review panel for their important contributions to the planning and direction of the study: Jomills Braddock, University of Miami; Adam Gamoran, University of Wisconsin at Madison; Tom Kane, Harvard University; Aaron Pallas, Columbia University; Kent Phillippe, American Association of Community Colleges; Russ Rumberger, University of California, Santa Barbara; Barbara Schneider, University of Chicago; Scott Swail, The College Board; Dawn Terkla, Tufts University; and Vince Tinto, Syracuse University.

The authors also wish to thank those who contributed to the production of this manual. The following individuals at RTI assisted in this effort: Laura Burns, Lil Clark, John Doherty, Saju Joshua, Lynne Kline, Shari Lambert, Lei Li, Jen Marenberg, Joe McMichael, Amy Sommer, and Milorad Stojanovic.

Finally, our special thanks to others who contributed to the success of this final round of NELS:88 data collection: Dan Pratt of RTI, who directed the initial stages of the fourth follow-up project, and Phil Kaufman and his colleagues at MPR Associates, who assisted in the work as RTI's subcontractor.

Table of Contents

List of Tables

List of Figures

Chapter I
Introduction to the National Education Longitudinal Study of 1988

This data file user's manual documents the procedures and methodologies employed during the National Education Longitudinal Study of 1988 (NELS:88). The manual is designed to provide guidance and documentation for users of the public-release data for the base-year data collection in 1988 through the fourth follow-up study, which took place 12 years later in 2000 (NELS:88/2000). Included in this report are also the results of the intermediate data collections, which took place in 1990 (the first follow-up study, 1992 (the second follow-up study), and 1994 (the third follow-up NELS:88/94). This manual will familiarize the user with each wave of NELS:88.

1.1 Overview of the Data User's Manual

NELS:88 data sets have been produced in both public- and restricted-use form. The ***public-use*** data files reflect alteration or suppression of some of the original data to minimize disclosure risk to individual respondents.[1] The ***restricted-use*** files preserve the original data, free of most confidentiality edits. Data files with high disclosure potential, specifically the high school and postsecondary education transcript files, are available in restricted form only. A more detailed discussion of measures used to preserve respondent confidentiality and procedures for gaining access to restricted-use data may be found in chapter 5 of this document.

This comprehensive user's manual includes information on the base-year data collection and all subsequent waves of NELS:88. It should be noted, however, that other, more detailed, data file user's manuals were developed for the base-year, first follow-up, and second follow-up studies. In addition, a detailed methodology report was prepared for the third follow-up study in 1994. These earlier documents include additional information on record layouts, codebooks, and questionnaires and variable descriptions for these prior waves. Information on these publications and on other available documentation for NELS:88 is discussed later in this chapter and in subsequent sections of this report.

This document is segmented into five chapters with supporting appendices that describe the study's data collection and processing procedures, sampling and weighting, and methodological outcomes for each wave of NELS:88. Chapter 1 presents information about the purposes of the study, its objectives, and the history of the National Center for Education Statistics' (NCES') education longitudinal studies program. Chapter 2 describes the data collection instruments used during each of the study waves, from the group-administered paper-and-pencil tests and questionnaires of the in-school base year and first two follow-ups, to the computer-assisted interviews used for the two post-high school follow-ups. The complex sample design and weighting for this longitudinal study are presented in chapter 3. Data collection and data processing procedures for the five data collection waves are described in chapter 4 and chapter 5, respectively. Appended to this report are materials designed to

[1] For example, knowing unique identifiers for a sample member's junior or senior high school and the postsecondary educational institutions he or she attended could potentially identify the individual.

provide additional documentation and support to the NELS:88 data user. These appendices include, among other materials, a study glossary, summaries of response rates, further tables on fourth follow-up standard errors and design effects, and information on fourth follow-up (2000) derived variables.

In the presentation that follows, discussion of the in-school rounds (1988-1992) is to a large extent consolidated and presented as a block. This discussion is then followed by separate presentations for the two out-of-school rounds in 1994 and 2000. This reflects the structure of the three Electronic Codebooks (ECBs) that cover these periods.[2] For example, the 1988-92 public-use ECB (N2P) includes for analysis all cases (n = 27,394) with respondents from any of the first three rounds or sample member cohorts (8^{th}-, 10^{th}-, and 12^{th}-grade students). This ECB also includes students who were originally excluded from the study but whose eligibility status changed over time. The 1994 ECB also contains data for the first three rounds, but it is based on the 1994 sample, with its 14,915 respondents.[3] The third NELS:88 ECB (N0P) contains data for all five rounds, but is based on the 2000 sample with its 12,144 respondents.[4] This data partitioning is further supported by the similarity of the in-school rounds in methodology and design, and the dissimilarity of the out-of-school rounds to earlier in-school data collections. The in-school rounds have a common data collection methodology (achievement tests and questionnaires were administered in paper-and-pencil formats and optically scanned) and include contextual components such as surveys of parents, teachers, and principals. The two out-of-school rounds primarily used computer-assisted telephone interviewing (CATI) as the data collection mode, and included no contextual components.

1.2 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, the U.S. Department of Education's National Center for Education Statistics (NCES) instituted the National Education Longitudinal Studies program. The aim of this continuing program is to study the educational, vocational, and personal development of students at various stages in their educational careers, and the personal, familial, social, institutional, and cultural factors that may affect that development.

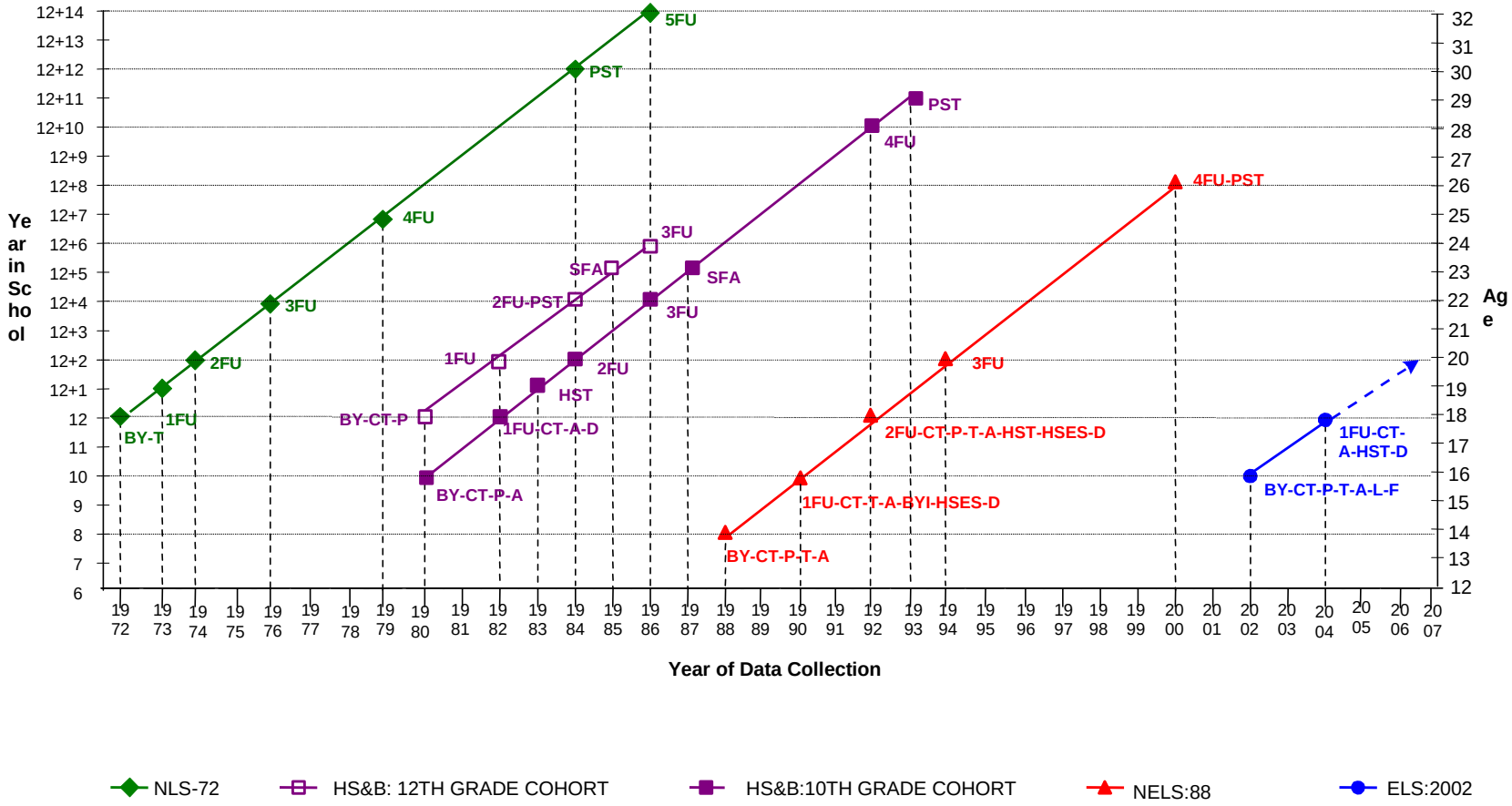
The high school longitudinal studies program consists of three completed studies: The National Longitudinal Study of the High School Class of 1972 (NLS-72), High School and Beyond (HS&B), and the National Education Longitudinal Study of 1988 (NELS:88). In addition, data collection for the Educational Longitudinal Study of 2002, the fourth longitudinal study in this time series, is currently in progress. Taken together, these studies describe (or will describe) the educational experiences of students from four decades—the 1970s, 1980s, 1990s, and 2000s—and not only describe and measure educational attainment but also provide bases for further understanding the correlates of educational success in the United States. Figure 1.2 includes a temporal presentation of these four longitudinal education studies, and highlights their component and comparison points.

[2] ECBs are electronic versions of fully documented survey codebooks developed by NCES. Users can browse through NELS:88 instruments, search variables and value labels, and output data for statistical analyses.

[3] These two ECBs (1988-1992, called N2P, and 1988-1994, called N4P) were released in March 1996.

[4] A fourth ECB containing data and documentation from the NELS:88/2000 postsecondary education transcript study is forthcoming.

Figure 1.2.--Longitudinal design for the NCES high school cohorts



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

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

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

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

1.2.1 Enabling Legislation

NCES and NELS:88 are authorized by Section 406(b) of the *General Education Provisions Act* (20 United States Code 1221e), as amended by the *National Education Statistics Act of 1994* (Public Law 103-382). This legislation requires that NCES, among other things:

- Collect, acquire, compile, and disseminate full and complete statistics on the condition and progress of education at the preschool, elementary, secondary, and postsecondary levels in the United States;
- Conduct and publish reports and analyses of the meaning and significance of such statistics; and
- Conduct longitudinal studies to report on the condition and progress of education (*National Education Statistics Act of 1994*, Section 404).

1.2.2 National Longitudinal Study of the High School Class of 1972

The Education Longitudinal Studies program began over 30 years ago, with the implementation of the National Longitudinal Study of the High School Class 1972 (NLS-72).[5] NLS-72 was designed to provide longitudinal data for educational policymakers and researchers that linked 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.

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, degree of satisfaction with education 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, and community involvement, and personal evaluations of educational activities were also included in the study.

1.2.3 High School and Beyond

Almost 10 years after the start of NLS-72, the second in the series of NCES longitudinal studies was launched. High School and Beyond (HS&B) included one cohort of high school seniors comparable

[5] Riccobono, J.A., Place, C., and Burkheimer, G.J. (1981). *National Longitudinal Study: Base Year through Fourth Follow-Up.* Washington, DC: U.S. Department of Education, National Center for Education Statistics; Tourangeau, R., et al. *The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up (1986) Data File User's Manual*, Washington, DC: U.S. Department of Education, National Center for Education Statistics.

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 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 responded in 1992. In addition, data were collected from teachers, principals, and parents to better understand the school and home contexts for the sample members. As in NLS-72, secondary and postsecondary transcripts were collected for the HS&B cohorts.

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 showing 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 cognitive tests within the longitudinal study to assess growth in cognitive abilities over time. Moreover, data were then available to analyze the school experiences of students who later dropped out of high school. 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.[6]

1.2.4 National Education Longitudinal Study of 1988

Much as NLS-72 captured a high school cohort of the 1970s and HS&B 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. Data collection for the National Education Longitudinal Study of 1988 was initiated with the 8th-grade class of 1988. At that time, NELS:88 was the most ambitious longitudinal study undertaken by NCES. It further extended the age and grade span of NCES longitudinal studies by collecting data from a middle school/junior high school cohort. Along with the student survey, NELS:88 included surveys of parents, teachers, and school administrators. By beginning with the 8th-grade, NELS:88 was able to capture the population of early dropouts—those who left school prior to spring term of 10th grade—as well as later dropouts (who left after spring of 10th grade) as had been studied in HS&B. The study was designed not only to follow a cohort of students over time (as had the predecessor studies) but also to "freshen" the sample at each of the first two follow-ups, and thus to follow multiple grade-defined cohorts over time. Thus, 10th grade and 12th grade cohorts were included in NELS:88 in the first follow-up (1990) and the second follow-up (1992), respectively. The freshening of the sample not only provided comparability to earlier cohorts from NLS-72 and HS&B, but it enabled researchers to conduct both grade representative cross-sectional and subsequent longitudinal analyses with the data. In late 1992 and early1993, high school transcripts were collected for sample members, and, in the fall of 2000 and early 2001, postsecondary transcripts were collected, further increasing the analytic potential of the data. Consequently, NELS:88 represents an integrated system of data that tracked students from middle school through secondary and postsecondary education, labor market experiences, and marriage and family formation.

Each of the five NELS:88 data collection rounds—1988 base year through year 2000 fourth follow-up—is described below. NELS:88 was conducted, in behalf of NCES, by two prime contractors. The base-year through third follow-up surveys were conducted by the National Opinion Research Center

[6] For a summary of reforms instituted between the time the HS&B cohort was in high school and the time the NELS:88 cohort was in middle/junior high and high school, see *America's High School Sophomores: A Ten Year Comparison* (Rasinski, K., Ingels, S.J., Rock, D.A., and Pollack, J.M., NCES 93-087), or *The Education Reform Decade* (Barton, P., and Coley, R., 1990, Princeton, NJ: Educational Testing Service).

(NORC) at The University of Chicago. The fourth follow-up survey was conducted by Research Triangle Institute (RTI) in North Carolina.

Base-Year Study. The base-year survey for NELS:88 was carried out during the 1988 spring semester. The study employed a clustered, stratified national probability sample of 1,052 public and private 8th-grade schools. Almost 25,000 students across the United States participated in the base-year study. Questionnaires and cognitive tests were administered to each student in the NELS:88 base year. The student questionnaire covered school experiences, activities, attitudes, plans, selected background characteristics, and language proficiency. School principals completed a questionnaire about the school; two teachers of each student were asked to answer questions about the student, about themselves, and about their school; and one parent of each student was surveyed regarding family characteristics and student activities.

First Follow-up Study. Conducted in 1990, when most sample members were high school sophomores, the first follow-up included the same components as the base-year study, with the exception of the parent survey. Additionally, a "freshened" sample was added to the student component to achieve a representative sample of the nation's sophomores; thus, trend comparisons were made possible between the 1980 HS&B sophomore cohort and NELS:88 1990 sophomores. The study frame included 19,363 in-school students, and 18,221 sample members responded. Importantly, the first follow-up study tracked base-year sample members who had dropped out of school, with 1,043 dropouts taking part in the study. Overall, the study included a total of 19,264 participating students and dropouts. In addition, 1,291 principals took part in the study, as did nearly 10,000 teachers.

Second Follow-up Study. The second follow-up took place early in 1992, when most sample members were in the second semester of their senior year. The study provided a culminating measurement of learning in the course of secondary school and also collected information that facilitated the investigation of the transition into the labor force and postsecondary education. Because the NELS:88 longitudinal sample was freshened to represent 1992 seniors, trend comparisons were made possible to the senior cohorts of NLS-72 and HS&B. The NELS:88 second follow-up resurveyed students who were identified as dropouts in 1990, and identified and surveyed additional students who had left school since the previous wave. For selected subsamples, data collection also included the sample member's parents, teachers, school administrators, and academic transcripts.

Third Follow-up Study (NELS:88/94). The NELS:88 third follow-up took place early in 1994. By this time in their educational careers, most of the sample members had already graduated from high school, and many had begun postsecondary education or entered the workforce. The study addressed issues of employment and postsecondary access and was designed to allow continuing trend comparisons with other NCES longitudinal studies. For the first time in the sequence of NELS:88 studies, the primary form of data collection was individual computer-assisted telephone interviews (CATI), with personal interviews with selected respondents who required intensive tracking and nonresponse refusal conversion.

Fourth Follow-up Study (NELS:88/2000). The fourth follow-up to NELS:88 (NELS:88/2000) included interviews with 12,144 members of the three NELS:88 sample cohorts 12 years after the base-year data collection. Because these data represent the period 6 years after the last contact with the sample, they will enable researchers to explore a new set of educational and social issues[7] about the NELS:88 respondents. For example, in 2000, most of the participants from the various cohorts of

[7] For an extended discussion of research questions that can be addressed with NELS:88 fourth follow-up data, see appendix B of Ingels, Curtin, Kaufman, Alt and Chen, 2002, *Coming of Age in the 1990s: The Eighth-Grade Class of 1988 12 Years Later*, NCES 2002-321.

NELS:88 had been out of high school for 8 years and were 26 years old. At this age, the majority of students who intend to enroll in postsecondary schools will already have done so. Thus, a large proportion of students have completed college; some completed graduate programs. Many of these young people have married and have children of their own; some are divorced; some are successful in the market place, while others have had less smooth transitions into the labor force.

As in the third follow-up, data collection for the NELS:88 fourth follow-up was primarily by CATI; however, laptop-based computer-assisted personal interviews (CAPI) were used in situations where more intensive field follow-up was required. The study focused on postsecondary education and employment, and especially the transitions experienced by young adults as they moved from educational systems (secondary and postsecondary) into the labor market. Interview topics included experiences with postsecondary education, labor market outcomes, job-related training, community integration, and marriage and family formation. The fourth follow-up study also collected transcripts from the postsecondary institutions that study respondents reported attending after high school.

1.2.5 Education Longitudinal Study of 2002 (ELS:2002)

The Education Longitudinal Study of 2002 is the fourth in the NCES national longitudinal high school cohort series and is designed to build on the multiple policy objectives of NLS-72, HS&B, and NELS:88. Base-year data collection for the study is currently (April, 2002) taking place, with approximately 20,000 10th grade students selected from 750 public and private high schools. Policy issues to be studied through ELS:2002 include the identification of school attributes associated with achievement; the influence of parent and community involvement on students' achievement; the dynamics and determinants of dropping out of the educational system; changes in educational practices over time; and the transition of different racial-ethnic, gender, and socioeconomic groups from high school to postsecondary institutions and the labor market. Like the earlier studies, ELS:2002 will examine students' values and goals, investigate factors affecting risk and resiliency, and gather information about participation in social and community activities. The study will also obtain teacher evaluations of the effort and ability of each student, school administrator questionnaires, school library and media center questionnaires, and parent questionnaires. A facilities checklist will be completed for each school, noting features of its physical plant and degree of order and safety. In the ELS:2002 first follow-up (2004), high school transcripts will be collected covering the span of the high school years.

As in NELS:88, ELS:2002 will include measures of school climate, each student's native language and language use, student and parental educational expectations, attendance at school, course and program selection, use of technology, planning for college, interactions with teachers and peers, perceptions of safety in school, parental income, resources, and home education support system.

The longitudinal study is also designed to support both longitudinal and cross-cohort analyses and to provide a basis for important descriptive cross-sectional analyses as well. However, priority was given to the longitudinal aspects of the study, with survey items chosen for their usefulness in predicting or explaining future outcomes as measured in later survey waves. The ELS:2002 content is also designed to provide comparability, where possible, to the prior NCES high school studies to facilitate cross-cohort comparisons. For example, trends over time can be examined by comparing the data from 1980, 1990, and 2002 high school sophomores, collected with HS&B, NELS:88, and ELS:2002, respectively; or data from 1972, 1980, 1982,[8] 1992, and 2004 high school seniors, collected from NLS-72, HS&B, NELS:88, and ELS:2002.

[8] Because the HS&B 1980 sophomore cohort sample in 1982 was not freshened, it constitutes only an approximation of a senior year sample in that it represents only 1982 seniors who were sophomores 2 years before.

1.2.6 Other Related NCES Studies

In addition to the high school cohort longitudinal studies just discussed, other data collection activities sponsored by NCES are also relevant to users of the NELS:88 data. For example, these program areas include other studies of educational performance and outcomes, such as longitudinal studies with older, college-aged students, as well as longitudinal birth and kindergarten cohorts. In addition, cross sectional time series assessments and surveys of students, educational institutions and educational personnel also complement the research objectives of NELS:88. We provide brief descriptions of these programs below.

Other NCES Longitudinal Studies. The ***Early Childhood Longitudinal Study (ECLS) Program*** comprises two studies, a 1998 kindergarten cohort (ECLS-K) and birth cohort of children born in 2001 (ECLS-B). The birth cohort will follow a sample of children through the 1st grade; the kindergarten cohort will follow a sample of children from kindergarten through the 5th grade. The ECLS program has both descriptive and analytic purposes. The descriptive purposes are to provide national data on children's status at birth and at various points thereafter; children's transitions to nonparental care, early education programs, and school; and children's experiences and growth through the 5th grade. The analytic purpose of the program is to provide data to test hypotheses about the effects of a wide range of family, school, community, and individual variables on children's development, early learning, and early performance in school.

The ***National Postsecondary Student Aid Study (NPSAS)*** is a comprehensive nationwide study of how students and their families pay for postsecondary education. It covers national representative samples of undergraduates, graduates, and first-professional students, including students attending less than 2-year institutions, 2- to 3-year schools, 4-year colleges, and major universities. Participants include students who do not receive aid and their parents, as well as students who do receive financial aid and their parents. Study results are used to help determine future federal policy regarding student financial aid. The study has been conducted every 3 years since the 1986-87 school year, when 1,074 postsecondary institutions, 60,000 students, and 14,000 parents responded. Beginning in 1999-2000, the survey will be conducted every 4 years. The NPSAS sample is used to supply the baseline for two postsecondary longitudinal studies, described below.

Drawing students from the NPSAS sample, the ***Beginning Postsecondary Students Longitudinal Study (BPS)*** provides information on persistence, progress, and attainment from initial time of entry into postsecondary education through leaving and entering the workforce. BPS allows us to study what happens to a cohort of incoming college freshman. Unlike NELS:88, it includes nontraditional students (e.g., late entrants into the postsecondary sector), as well as traditional students (i.e., those who go directly from high school to the postsecondary sector, or start college within a year or two of graduation). BPS follows first-time, beginning students for at 5 five years at approximately 2-year intervals, collecting student data, and financial aid reports. By starting with a cohort that has already entered postsecondary education and following it for 5 years, BPS is able to determine to what extent students who start postsecondary education at various ages differ in their progress, persistence, and attainment. The first BPS study was conducted in 1989-90, with follow-up surveys in 1992 and 1994. The second BPS cohort of students began with a survey in 1995-96 and included follow-ups in 1998 and 2001.

Like the BPS, the ***Baccalaureate and Beyond Longitudinal Study (B&B)*** also uses NPSAS participating students as its base-year sample (for example, NPSAS:93 "spun off" B&B:93/94; and NPSAS:96 spun off BPS:96/98). The study provides information concerning education and work experience for students after completing their bachelor's degrees. B&B provides cross-sectional information 1 year after bachelor's degree completion, while at the same time providing longitudinal data concerning entry into, persistence and progress through, and completion of graduate-level education.

Such information is not available through follow-ups involving high school cohorts or even college-entry cohorts, both of which are restricted in the number who actually complete a bachelor's degree and continue their education, and limited also in the number of years respondents are followed.

About 11,000 students who completed their degrees in the 1992-93 academic year were included in the first B&B (B&B:93/94). In addition to the student data, B&B collected postsecondary transcripts covering the undergraduate period, which provided complete information on progress and persistence at both the undergraduate and graduate levels. The second B&B follow-up took place in spring 1997 (B&B:93/97) and gathered information on employment history, family formation, and enrollment in graduate programs. A third follow-up of the 1993 B&B cohort will take place in spring 2003. A new cohort, B&B:2000/2001, is currently underway.

Schools and Staffing Survey. The Schools and Staffing Survey (SASS) collects sample survey data on American public and private elementary and secondary schools. Initiated in 1987-88 as a comprehensive study to identify and describe potential shortages in the public and private school work force, the study is designed to address the critical need for data on the characteristics of teachers and administrators and the conditions of their working environments. SASS is a comprehensive, public and private, and elementary and secondary, education database that combines and expands on 12 separate surveys that NCES has conducted in the past, including surveys of teacher supply and demand, public and private schools, public and private school teachers, public and private school administrators, students, public and private school libraries and media centers, and public and private school librarians.

First conducted during the 1987-88 school year and subsequently repeated during 1990-91, 1993-94, and 1999-2000 school years, the study design includes schools as the primary sampling unit and then samples teachers from each selected school; public school districts are included in the sample when one or more of their schools are selected. SASS also encompasses a Teacher Follow–up Survey the purpose of which is to determine how many teachers remained at the same school, moved to another school or left the profession in the year following the SASS administration.

SASS will be next offered during the 2003-04 school year. During the 1999-2000 SASS, estimates were based on the responses of a sample of approximately 9,900 public schools, 3,500 private schools, and 5,500 public school districts associated with the public schools in the sample. From these schools, about 56,500 public and 11,000 private school teachers were selected. The 1999-2000 SASS also included 13,600 school libraries and media centers and a sample of 1,100 charter schools and 4,400 teachers in those schools.

National Assessment of Educational Progress. The National Assessment of Educational Progress (NAEP) is a series of cross-sectional studies first implemented in 1969 to collect information about educational achievement in the United States. Since implementation, NAEP assessments have included U.S. students partitioned both by age (9-, 13-, and 17-year-olds and young adults aged 25 to 35 years) and grade (4^{th}, 8^{th}, and 12^{th} grades), and learning area (to date, 10 areas including mathematics and English). NAEP administers national and state-level assessments (Main NAEP) as well as assessments of long-term trends in educational achievement (Trend NAEP).

Trend NAEP is designed to give information on changes in the basic achievement of U.S. youth. Nationally representative samples of students have been assessed in science, mathematics, and reading at ages 9, 13, and 17 since the early 1970s; and assessed in writing in grades 4, 8, and 12 since 1984. To measure trends accurately, assessment items (mostly multiple choice) and procedures have remained unchanged since the first assessment in each subject. Recent trend assessments were conducted in 1994, 1996, and 1999; about 30,000 students took part in the 1996 trend assessment. Using the 1992 NAEP assessment in mathematics, NELS:88 1992 mathematics test results were converted to the NAEP

reporting scale (and proficiency levels). Because NAEP and NELS:88 are vertically scaled, the 1992 crosswalk also serves to place the NELS:88 1988 and 1990 mathematics results on the NAEP scale.

Since 1990, the Main NAEP assessments of achievement at the national level have also been conducted at the state level for some states. Participating states receive assessment results that describe the performance of students in that state. In its content, the state assessment is identical to the assessment conducted nationally. However, because the national NAEP samples were not and are not currently designed to support the reporting of accurate and representative state-level results, separate representative samples of students are selected for participating states. In 1998, the Main NAEP assessed reading in the 4th and 8th grades and writing in the 8th grade; the 2000 assessments included math and science assessments in the fourth and 8th grades.

In addition to the assessments discussed above, NAEP also coordinates a number of other educational studies related to assessment. Ongoing projects include high school transcript studies and a technology-based assessment project designed to explore the use of technology in schools, especially computer use to enhance the quality and efficiency of educational assessments. The HS&B, NELS:88 and NAEP transcript studies have employed a common course coding system and methodology so that together they provide a time series for high school seniors from 1982 through 2000.

1.3 NELS:88 Study Objectives

The major features of NELS:88 included the planned integration of data from students, dropouts, parents, teachers, and schools, with an initial focus on an 8th grade cohort and follow-up studies taking place at 2-year intervals. This core design was supplemented to gather data on other key research areas, including high school and postsecondary transcript studies, and a high school effectiveness study. Multiple research and policy objectives are addressed through this design. The study is intended to support a general purpose dataset for the development and examination of federal educational policy. Part of its aim is to inform decisionmakers, educational practitioners, and parents about the changes in the operation of the education system over time and the effects of various elements of the system on the lives of the individuals who pass through it. Specifically, NELS:88 focuses on a number of interrelated policy issues, including identification of school attributes associated with achievement; the transition of different types of students from 8th grade to secondary school and to postsecondary institutions; the transition of secondary and postsecondary students to the workforce; the influence of ability grouping and program type on future educational experiences and achievements; determinants of dropping out of the education system; and changes in educational practices over time. One of the defining features of NELS:88 is the extensive attention it gives to the role of parents. For example, parent questionnaires were completed during the base-year and second follow-up studies and collected useful information on parents' attitudes and behaviors on educational or career choices, financial preparation for postsecondary education, the correlates of parental involvement in schools, and the parents' role in the educational success of their children. Information is also included on residential neighborhoods (some of this information, such as residential zip code mappings to 1990 Census community variables or zip codes in 2000, is available only on a restricted use basis).

The NELS:88 design enables researchers to conduct analyses on three levels: crosswave (by following a single group of individuals as they develop over time), cross-sectional (at a single time point), and cross-cohort (by comparing NELS:88 findings to those of HS&B and NLS-72). The first of these levels provides NELS:88 with its primary objective: to serve the purposes of longitudinal measurement. The sampling and data collection designs give priority to maintaining and surveying a substantial number of base-year sample members, as well as to sustaining overlapping but analytically distinct cohorts of sophomores and seniors. Users of NELS:88 data will be able to study the effects of a wide variety of

factors on students' educational and professional attainment. The longitudinal data gathered from students and augmented through parent, teacher, school administrator, and school record (i.e., high school and postsecondary transcripts) accounts of student's progression and development, will facilitate scrutiny of various facets of student's lives—their problems, successes and concerns; their relationships with parents, peers, and teachers; and the characteristics of their schools—and permit examination of the impact of these factors on social, behavioral, and educational development.

The second analytic level within NELS:88 is cross-sectional. By beginning with a cross-section of 1988 8th-graders, following a substantial subsample of these students at 2-year intervals, and freshening the 1990 and 1992 samples to obtain representative national cross-sections of 10th- and 12th-graders, the study also provides a statistical profile of America's 8th-graders, high school sophomores, and high school seniors, as well as a profile of both early and late high school dropouts.

Finally, NELS:88 has been designed to provide researchers with data for drawing comparisons with previous NCES longitudinal studies (as well as comparisons with future longitudinal efforts). After the release of the NELS:88 first follow-up data, researchers were able to conduct trend analyses with the 1980 sophomore cohort of HS&B. With completion of the NELS:88 second follow-up, NELS:88, HS&B, and NLS-72 senior cohort comparisons became possible. To facilitate cross-cohort comparisons, many of the questions contained in the prior survey were repeated in NELS:88, and data processing and file conventions were kept consistent, to the extent possible, with HS&B and NLS-72. For users specifically interested in conducting trend analyses of NLS-72, HS&B, and NELS:88 data, further information on content and design similarities and differences is presented later in this report.

Table 1.3 provides an overview of the components for each of the five waves of NELS:88. The study designs for each of the waves are described in the following sections.

1.3.1 Base-year Study and Sample Design

The base-year study design consisted of four components: survey and tests of students, and surveys of parents, school administrators, and teachers. A student questionnaire gathered information about basic background variables and a range of other topics, including school work, educational and occupational aspirations, and social relationships. Students also completed a series of curriculum-sensitive cognitive tests to measure educational achievement and cognitive growth between 8th and 12th grades in four subject areas: reading, mathematics, science, and social studies. One parent of each student was asked to respond to a parent survey intended to measure parental aspirations for children, family willingness to commit resources to children's education, the home educational support system, and other family characteristics relevant to achievement. Selected teachers in two of the four subject areas completed a teacher questionnaire designed to collect data about school and teacher characteristics, evaluations of the selected students, course content, and classroom teacher practices. Finally, school principals completed school administrator questionnaires. This administrator questionnaire gathered descriptive information about the school's teaching staff, the school climate, characteristics of the student body, and school policies and offerings.

Table 1.3.—Data sources for the National Education Longitudinal Study of 1988, by year and data collection wave: 1988-2000

Data Collection (Year)	Data Source					
	Students/ Dropouts[1]	Parents	School Administrators	Teachers	Cognitive Tests	Transcripts[2]
Base Year	X	X	X	X	X	
First Follow-up	X		X	X	X	
Second Follow-up	X	X	X	X	X	X
Third Follow-up	X					
Fourth Follow-up	X					X

[1] Following the base-year data collection in 1988, all NELS:88 follow-up studies included school dropouts.

[2] Data collection in 1992 included high school transcripts; postsecondary transcripts were collected in 2000.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988, fourth follow-up study (NELS:88/2000), 1988-2000.

In the NELS:88 base year, a two-stage probability design was used to select a nationally representative sample of 8th grade school and students. Schools constituted the primary sampling unit; the target sample size for schools was 1,032. A pool of 1,032 schools was selected through stratified sampling with probability of selection proportional to 8th grade size and with oversampling of private schools. A pool of l,032 replacement schools was selected by the same method. Of the 1,032 initial selections, 30 proved to be ineligible. Of the 1,002 eligible selections, 698 participated. An additional 359 schools supplied by alternative selections available from the replacement pool also participated, for a total school sample of 1,057 cooperating schools, of which 1,052 schools (815 public schools and 237 private schools) contributed usable student data. For 1,035 of these 1,052 schools, both student and school administrator data were received. In the NELS:88 base-year design, students were the second stage sampling unit. A random selection of 26,432 students from participating sampled schools resulted in participation by 24,599 spring term 1988 8th-graders. On average, 23 student participants represented each of the participating schools. Additional information about the base-year sample design is provided in the *NELS:88 Base Year Sample Design Report;*[9] see also chapter 3 of this manual.

1.3.2 First Follow-up Core Study and Sample Design

The first follow-up of NELS:88 comprised the same components as the base-year study, with the exception of the parent survey, which was not repeated in the 1990 round. In addition, three new components—the dropout study, base-year ineligible study, and high school effectiveness study—were initiated in the first follow-up, and a freshened sample was added to the student component. As in the base-year, students were asked to complete a questionnaire and cognitive test. The cognitive test was designed to measure 10th-grade achievement and cognitive growth between 1988 and 1990 in mathematics, science, reading, and social studies (history/geography/civics). The student questionnaire

[9] Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A., and Tourangeau, R.E. (1990*). NELS:88 Base Year Sample Design Report.* (NCES 90-463). Washington, DC: U.S. Department of Education.

collected basic background information and asked students about such topics as their school and home environments, participation in classes and extracurricular activities, current jobs, their goals and aspirations, and opinions about themselves. Following the base-year design, two teachers of each student were asked to complete a teacher questionnaire, and school principals completed a school administrator questionnaire. First-time participants in NELS:88—including students just added to the cohort through the sample freshening process, base-year ineligibles who became eligible in the first follow-up, and base-year nonrespondents who did participate in the first follow-up—completed a new student supplement, containing basic demographic items, which were asked in the base year but not repeated in the first follow-up. The first follow-up also surveyed and tested youths who had dropped out of school at some point between the spring term of the 1987-88 school year and that of the 1989-90 school year. The dropout questionnaire collected information on a wide range of subjects, including reasons for leaving school, school experiences, absenteeism, family formation, plans for the future, employment, attitudes and self-concept, and home environment.

Student sample selection was implemented in two stages. In the first stage, 21,474 students who were in the 8th grade NELS:88 sample in 1988 were selected. In addition, because some sophomores in 1990 were either not in the country or not in the 8th grade during 1988 (when base-year data collection took place), the sophomore cohort was augmented through a process called "freshening." The procedure was designed to provide a representative sample of students enrolled in the 10th grade in the 1989-90 school year, comparable to the HS&B sophomore cohort. Freshening added an additional 1,034 high schools students who were not contained in the base-year sampling frame, and 855 were considered eligible. (Further information on the first follow-up sample design can be found in chapter 3, 3.4.1).

Several components were added to the first follow-up to increase its analytic power. One of these enhancements, the base-year ineligible (BYI) study, was added to the first follow-up in order to ascertain the 1990 school enrollment status and the 1990 NELS:88 eligibility status of students who were excluded from the base-year survey because of a language barrier or physical or mental disability that were thought to preclude them from completing the questionnaire and cognitive test. After the BYI, 341 students became eligible and were included in both the first and second follow-up studies.

In addition to the BYI study, the high school effectiveness study (HSES) was conducted in conjunction with the first follow-up study. The HSES was designed to allow NELS:88 analysts to better study school effects on education by augmenting the NELS:88 sample of urban and suburban schools. (This was a design similar to one employed with the HS&B sophomore cohort.) The within-school student sample of 248 participating first follow-up high schools in the 30 largest metropolitan statistical areas was augmented to produce a probability sample of both schools and students within the framework of the primary longitudinal study.

1.3.3 Second Follow-up Core Study and Sample Design

The NELS:88 second follow-up repeated all components of the first follow-up study. In addition, for a subsample of students, parent questionnaire data were collected (the first follow-up study did not have a parent questionnaire). The second follow-up parent questionnaire included a supplemental question series for parents new to NELS:88 (such as parents of 1990 and 1992 freshened students), so that critical information obtained in the base year parent questionnaire could be obtained. Two new components—the high school transcript and course offerings components—were initiated in the second follow-up study. The course offerings component was implemented as a part of the HSES. The high school transcript component was undertaken for sample members as described later in this chapter. The second follow-up also used sample freshening to achieve a representative sample of students enrolled in the 12th grade during the spring term of the 1991-92 school year. The freshening added 279 students who were not in the 8th grade during the Spring of 1988 or the 10th grade during the Spring of 1990.

As in the previous waves, students were asked to complete a questionnaire and a series of cognitive tests. The cognitive tests were designed to measure 12th -grade achievement and cognitive growth between 1988 and 1992 in mathematics, science, reading, and social studies. The student questionnaire asked students about such topics as academic achievement; their perceptions and feelings about their curriculum and school, family structure and environment; social relations; and aspirations, attitudes, and values, especially as they relate to high school and occupational or postsecondary educational plans. The student questionnaire also gathered data about the family decision-making structure during the critical transition from secondary school to postsecondary education or the work environment. The student questionnaire contained a supplement for early graduates, the intent of which was to document the reasons for and circumstances of early graduation.

In a departure from the base-year and first follow-up teacher survey designs, owing to funding constraints, only one teacher (either a mathematics or science teacher) of each student was asked to complete a teacher questionnaire.[10] School principals completed a school administrator questionnaire, as in the first follow-up. If a student was a first-time participant in NELS:88, he or she also completed a new student supplement, containing basic demographic items that were asked in the base year but not repeated in the second follow-up student questionnaire.

The second follow-up, in addition to surveying students who were enrolled in school, surveyed and tested youths who had dropped out of school at some point between the spring term of the 1987-88 school year and the spring term of the 1991-92 school year. (Early graduates were also surveyed, as well as individuals who had completed high school by alternative means, such as exam certification [e.g., the GED].) The dropout questionnaire collected information on a wide range of subjects, including reasons for leaving school, school experiences, absenteeism, plans for the future, employment, attitudes and self-concept, and home environment.

Each student and dropout selected for the first follow-up study was selected with certainty in the second follow-up. From within the schools attended by the sample members (2,258), 1,500 12th-grade schools were selected as sampled schools. Of the 1,500 sampled schools, teacher surveys occurred in 1,374 schools. For students attending schools other than those 1,374 schools, only the student and parent questionnaires were administered.

The student sample was then augmented through freshening at the NELS:88 selected schools, the aim of which was to provide a representative sample of students enrolled in the 12th grade during the spring term of the 1991-92 school year. Freshening added an additional 364 12th-graders (of whom 279 were deemed eligible) who were not contained in either the base-year or first follow-up sampling frames.[11] The final sample size was 18,209 individuals eligible for the student component and 2,714 individuals eligible for the dropout component, or 20,923 in total. Additional information about the second follow-up sample design is provided in chapter 3 of this manual, with further detail in the *NELS:88 Second Follow-Up Student Component Data File User's Manual* (Ingels et al., NCES 94-374).

[10] If a student was not enrolled in either a mathematics or a science class, no teacher questionnaire was administered. During the spring of 1992, 10,861 students, 69.2 percent of the students in the contextual components sample, were enrolled in a mathematics class, a science class, or both.

[11] Of the 364 freshened students, 76 were sampling errors and became ineligible through questionnaire data; 15 dropped out of school between the sampling effort and data collection (these 15 are found only on the restricted-use file); 13 were out of scope because of a language barrier, moved out of the country, or were deceased; 9 were ineligible because of mental or physical incapacity; and the status could not be collected for 8 cases.

1.3.4 Second Follow-up Design Enhancements

Two new components, the transcript and the course offerings components, were added to the NELS:88 second follow-up. These components provide archival data that describe the academic experience of high school students and the curricula offered by their schools. The complete high school transcript record was collected for (1) the contextual sample—students attending sampled schools in the spring of 1992; (2) dropouts, dropouts in alternative programs, and early graduates, regardless of school affiliation; and (3) triple ineligibles[12] enrolled in the 12th grade in the spring of 1992, regardless of school affiliation. NELS:88 course-taking data provides a baseline against which future student outcome measures can be compared, but also illuminates trends when contrasted to the 1982 HS&B high school transcript study, NAEP transcript studies in 1987, 1990, 1994, 1998 and 2000. The course offerings component provides curriculum data from second follow-up high school effectiveness study (HSES) schools, through which school effects on student outcomes can be studied.

The high school effectiveness study (HSES) was added to the first follow-up to provide a probability sample of 10th-grade schools, with a sizable and representative within-school sample of students, through which longitudinal school-level analysis (comparable to 1980-82 HS&B sophomore cohort analysis) could be conducted. In the first follow-up HSES, permission to conduct the study was gained from 251 schools, and 248 of those schools were final HSES participants. The second follow-up HSES returned to 247 of the 251 cooperating first follow-up HSES schools and freshened both longitudinal and HSES sample members, and selecting additional students, including students who transferred to the school since the 1989 selection of HSES students. The second follow-up HSES was enhanced by the addition of archival data collected by the new course offerings component and was further augmented by the administration of free-response science and mathematics cognitive test items in HSES schools.

1.3.5 Third Follow-up Study and Sample Design

The NELS:88 third follow-up study (NELS:88/94) was designed to follow the ongoing progress of the NELS:88 sample cohorts as the members moved to a wide array of activities in postsecondary education and the world of work. NELS:88/94 examined issues of employment and postsecondary access, and sustained continuing comparisons with NLS-72 and HS&B. Specific content areas included academic achievement, perceptions and feelings about school and job, detailed work experiences, work-related training, and family structure and environment.

When the data were collected during the spring of the 1993-94 school year, it had been 2 years since most of the sample members had graduated from high school. However, also included in the group were students who had dropped out of school, or who had dropped out of school and subsequently returned. Hence, data collection for this diverse group departed from the school-based data collection used in earlier waves to employ computer-assisted telephone interviews (CATI), with field follow-up. The sample for the study was created by dividing the sample for the second follow-up study into 18 groups, based on their response history, dropout status, eligibility status, school sector types, race, test score, socioeconomic status, and freshened status. Each sampling group was assigned an overall selection probability. Cases within a group were selected such that the overall group probability was met but that the probability of selection within the group was proportional to each sample member's second follow-up weight. Haggerty, Dugoni, Reed, Cederlund, and Taylor (1996) describe this process in more detail. The final sample size was 15,875 individuals; subgroup breakdowns are reported in chapter 3, table 3.6-B of this report.

[12] Triple ineligibles are 1988 8th-graders who were ineligible for the base year, first follow-up, and second follow-up surveys because of mental or physical disability, or language barrier.

1.3.6 Fourth Follow-up Study and Sample Design

NELS:88/2000 is the fourth and final follow-up planned for the 8th grade class of 1988 (as well as the last contact with the 10th- and 12th-grade cohorts for the study). Data collection for the study involved a mixed mode approach: data were collected primarily through CATI but computer-assisted personal interviews (CAPI) were used with telephone nonrespondents. In addition to interviews with the sample members, the fourth follow-up also included a special postsecondary education transcript study for NELS:88 cohort members who responded to the fourth follow-up study and reported postsecondary education experience during either data collection for the third or fourth follow-up. Beginning at the end of CATI/CAPI data collection with the cohort members, the transcript study included requests to 3,213 postsecondary institutions, representing 16,020 postsecondary transcripts for the NELS:88 cohort.

In the course of the data collection year for this follow-up, most sample members had turned 26 years old. More than 6 years had passed since the last contact with them in 1994 when the NELS:88 third follow-up interview took place. These were important years for the sample members, many of whom had enrolled in postsecondary schools, started and changed careers, changed residences, and married and changed names. The sample member locating information from the previous follow-up interview was dated, requiring considerable resources to trace and locate sample members. (We discuss these tracing procedures in chapter 4 of this report.)

For cost reasons, and to limit the number of sample members who could not be located, the third follow-up frame was subsampled to limit the numbers of poor and difficult responders and sample members who were unlikely to be located (those who could not be located during earlier follow-up interviews). Data collection for the fourth follow-up study began in January, 2000, with a sample of 15,237 individuals, subsampled from the sampling frame of 15,964 individuals used for the third follow-up study in 1994, thus removing 647 third follow-up nonrespondents from the frame. Finally, at the end of data collection for the study, a second subsample of study nonrespondents took place to limit bias resulting from interview nonrespondents. The subsample included 386 fourth follow-up nonrespondents selected from American Indian and Alaska Native, Black non-Hispanic, Hispanic, and dropout strata.

1.4 Sources of Further Information on NELS:88

The following citations identify the technical reports and data file user's manuals produced for NELS:88. Information on many of these documents, as well as electronic links to some reports can be found on the NCES Web Site (http://nces.ed.gov/surveys/nels88/) or through the U.S. Government Printing Office (http://www.gpo.gov/). Most of these publications are also available through the ERIC system (http://www.askeric.org for the electronic catalogue to the ERIC database). In addition to these methodological documents, many research publications have been developed using NELS:88 data. For a comprehensive list of these publications, the interested reader is referred to the NELS:88 annotated bibliography on the NCES Web Site. This bibliography is also located on the public- and restricted-use ECBs (N0P/N0R) for the NELS:88 fourth follow-up study.

1.4.1 Base Year (1988)

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K., and Tourangeau, R. (1990). *NELS:88 Base Year Sample Design Report* (NCES 90-463). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Abraham, S., Karr, R., Spencer, B.D., and Frankel, M.R. (1990). *NELS:88 Base Year Student Component Data File User's Manual* (NCES 90-464). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Abraham, S., Rasinski, K.A., Karr, R., Spencer, B.D., and Frankel, M.R. (1990*). NELS:88 Base Year Parent Component Data File User's Manual* (NCES 90-466). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Abraham, S., Rasinski, K.A., Karr, R., Spencer, B.D., and Frankel, M.R. (1990*). NELS:88 Base Year School Component Data File User's Manual* (NCES 90-482). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Abraham, S., Rasinski, K.A., Karr, R., Spencer, B.D., and Frankel, M.R. (1990*). NELS:88 Base Year Teacher Component Data File User's Manual* (NCES 90-484). Washington, DC: National Center for Education Statistics.

Rock, D.A., and Pollack, J.M. (1991). *Psychometric Report for the NELS:88 Base Year Test Battery* (NCES 91-468). Washington, DC: National Center for Education Statistics.

Kaufman, P., and Rasinski, K.A. (1991). *Quality of the Responses of Eighth-Grade Students in NELS:88* (NCES 91-487). Washington, DC: National Center for Education Statistics.

1.4.2 First Follow-up (1990)

Ingels, S.J., Scott, L.A., Lindmark, J.T., Frankel, M.R., and Myers, S.L. (1992). *NELS:88 First Follow-Up Dropout Component Data File User's Manual* (NCES 92-083). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., Lindmark, J.T., Frankel, M.R., and Myers, S.L. (1992). *NELS:88 First Follow-Up School Component Data File User's Manual (NCES* 92-084). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., Lindmark, J.T., Frankel, M.R., and Myers, S.L. (1992). *NELS:88 First Follow-Up Student Component Data File User's Manual* (NCES 92-030). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., Lindmark, J.T., Frankel, M.R., and Myers, S.L. (1992). *NELS:88 First Follow-Up: Teacher Component Data File User's Manual* (NCES 92-085). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M., and Rasinski, K.A. (1994). *NELS:88 First Follow-Up Final Technical Report* (NCES 94-632). Washington, DC: National Center for Education Statistics.

1.4.3 Second Follow-up (1992)

Ingels, S.J., Dowd, K.L., Baldridge, J.D., Stipe, J.L., Bartot, V.H., and Frankel, M.R. (1994). *NELS:88 Second Follow-Up: Student Component Data File User's Manual* (NCES 93-374). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Dowd, K.L., Stipe, J.L., Baldridge, J.D., Bartot, V.H., and Frankel, M.R. 1994). *NELS:88 Second Follow-Up: Dropout Component Data File User's Manual* (NCES 93-375). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Thalji, L., Pulliam, P.A., Bartot, V.H., and Frankel, M.R. (1994). *NELS:88 Second Follow-Up: School Component Data File User's Manual* (NCES 94-376). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Dowd, K.L., Taylor, J.R., Bartot, V.H., and Frankel, M.R. (1994). *NELS:88 Second Follow-Up: Transcript Component Data File User's Manual* (NCES 94-377). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Thalji, L., Pulliam, P.A., Bartot, V.H., and Frankel, M.R. (1994). *NELS:88 Second Follow-Up: Parent Component Data File User's Manual* (NCES 94-378). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Thalji, L., Pulliam, P.A., Bartot, V.H., and Frankel, M.R. (1994). *NELS:88 Second Follow-Up: Teacher Component Data File User's Manual* (NCES 94-379). Washington, DC: National Center for Education Statistics.

Rock, D.A., and Pollack, J.M. (1995). *NELS:88 Base Year through Second Follow-Up Psychometric Report* (NCES 95-382). Washington, DC: National Center for Education Statistics.

Ingels, S.J. (1996). *Sample Exclusion in NELS:88–Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years* (NCES 96-723). Washington, DC: National Center for Education Statistics.

Scott, L.A., Ingels, S.J., Pulliam, P., Sehra, S., Taylor, J.R., and Jergovic, D. (1996*). NELS:88 High School Effectiveness Study: Data File User's Manual.* Chicago: NORC report to National Center for Education Statistics.

McLaughlin, D., and Cohen, J. (1997). *NELS:88 Survey Item Evaluation Report* (NCES 97-052). Washington, DC: National Center for Education Statistics.

Pollack, J.M., and Rock, D.A. (1997). *Constructed Response Tests in the NELS:88 High School Effectiveness Study* (NCES 97-804). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., and Taylor, J.R. (1998*). NELS:88 Base Year Through Second Follow-Up Final Methodology Report.* NCES Working Paper Series, 98-06. Washington, DC: National Center for Education Statistics.

1.4.4 Third Follow-up (1994)

Haggerty, C., Dugoni, B.L., Reed, L., Cederlund, A., and Taylor, J.R. (1996). *Methodology Report: National Education Longitudinal Study: 1988-1994* (NCES 96-174). Washington, DC: U.S Department of Education, National Center for Education Statistics.

Sanderson, A., Dugoni, B.L., Rasinski, K., and Taylor, J.R. (1996). *National Education Longitudinal Study 1988-1994 Descriptive Summary Report With an Essay on Access and Choice in Postsecondary Education*. (NCES 96-175). Washington, DC: U.S Department of Education, National Center for Education Statistics.

1.4.5 Fourth Follow-up (2000)

Ingels, S.J., Curtin, T.R., Kaufman, P., Alt, M.N., and Chen, X. (2002). *Coming of Age in the 1990s: The Eighth-Grade Class of 1988 12 Years Later*. (NCES 2002–321). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Chapter II
Data Collection Instruments

This chapter provides a brief description of the form and content of the student, new student supplement, dropout, school administrator, teacher, and parent survey instruments and cognitive tests used in the base year and first and second follow-ups for the NELS:88. It also describes the instruments used for all student and dropout sample members for the third and fourth follow-up interviews, which were conducted out of school, primarily by computer-assisted interview. In addition, this chapter provides information on the high school transcript component of the second follow-up and the postsecondary education transcript component of the fourth follow-up.

2.1 Overview of Instrument Development

With each new wave of the NELS:88 data collection, the research team enhanced the data collection instruments and added new instruments, striving to maintain similar content and form among instruments for the three in-school waves, while addressing at the same time new data elements appropriate for the age and experiences of the sample cohort. Instruments for the base year included a student questionnaire, student cognitive tests, and parent, teacher, and school administrator questionnaires. In the first and second follow-ups, interviewers re-administered these instruments, except for the parent questionnaire, and also added a dropout questionnaire for sample members who had left school and a new student supplement for students who were new to the sample (e.g., "freshened" at the first follow-up). The second follow-up then reintroduced the parent questionnaire—in revised form—and added a high school transcript component. Table 2.1 summarizes the instrumentation for the three in-school waves of NELS:88. A fuller account of the instrument development process may be found in appendix A.

In designing the NELS:88 questionnaires, the research team kept in mind the longitudinal goals of the study and chose items that would be useful in predicting or explaining outcomes captured in later survey waves. Team members also sought, on the one hand, to ensure continuity and consistency with earlier NCES education longitudinal studies, and on the other, to address new areas of policy concern and recent directions in theory. Where appropriate, they drew test and questionnaire content from NLS-72, HS&B, and other NCES studies, such as the National Assessment of Educational Progress (NAEP), the Second International Math Study (SIMS), and the Schools and Staffing Survey (SASS), to ensure a common standard of measurement that would permit comparisons with other important data sources and maximize the utility of NELS:88 data. For example, they designed the mathematics tests to allow test score comparisons with both the NAEP and HS&B. Readers interested in the crosswalks between the NELS:88 questionnaires and the HS&B and NLS-72 instruments are encouraged to refer to the data file user's manuals for the waves and components of interest.

One year before each wave of the NELS:88 main study, the research team field-tested data collection procedures and instruments. Thus, they conducted the first field test with the 8th grade class of 1987. They then used field test results to inform planning for the main study, improve the measurement properties of test and questionnaire items, and identify items that needed to be modified or deleted to improve the instrument length or item format.

Table 2.1.—NELS:88 school-based survey instruments, by wave of administration: 1988-1992

Survey Instrument	Survey Wave		
	Base Year	First Follow-up	Second Follow-up
Student questionnaire	Yes	Yes	Yes
Early graduate supplement	No	No	Yes
New student supplement	No	Yes	Yes
Dropout questionnaire	No	Yes	Yes
School administrator questionnaire	Yes[1]	Yes	Yes
Teacher questionnaire	Yes	Yes	Yes
Parent questionnaire	Yes	No	Yes
High school transcript component	No	No	Yes[2]

[1] In the base year, there were two school administrator surveys: in the spring of 1988, the regular NELS:88 principal survey, and in the fall of 1989, a special principal survey on the topic of middle grades practices.

[2] The high school transcripts, which were collected in the second follow-up, span the entire high school career, including 10th grade—the modal grade of first follow-up sample members—and typically 9th grade, as well.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

2.2. Base-year through Second Follow-up Student Questionnaires

For the base year, all sample members completed a student questionnaire. For the first and second follow-ups, NELS:88 project staff re-administered the student questionnaire to all sample members who were enrolled in school during the spring term of the survey year (spring 1990 for the first follow-up, and spring 1992 for the second follow-up). In the second follow-up, they also administered it to sample members who had left school and had completed the dropout questionnaire during the first follow-up but had since passed the General Educational Development (GED) test or obtained some other equivalency certification. Sample members completed the surveys at either in-school or off-campus survey sessions. Although the base-year questionnaire was only available in English, the first and second follow-up questionnaires were available in both English and Spanish.[13]

The 60-minute, self-administered student questionnaire used in each wave collected information on a wide range of topics, including

- student background
- language use

[13] Excluding the base-year ineligible students who were reclassified as eligible in the first follow-up, 19 students completed the Spanish-language questionnaire in the NELS:88 first follow-up. Eight dropouts and 41 students completed the Spanish-language questionnaire in the second follow-up. Because of the small numbers of questionnaires completed in Spanish, flags were not created to identify these cases. The percentage of questionnaires completed in Spanish in 1990 and 1992 is similar to the percentage of HS&B respondents who opted to complete Spanish-language questionnaires in 1980 and 1982. For copies of the Spanish-language questionnaires, see the technical reports for the first- and second follow-ups (Ingels et al., NCES 94-632 and NCES 98-06).

- home environment
- perceptions of self
- occupational or postsecondary educational plans
- jobs and household chores
- school experiences and activities
- work and social activities

Information collected in the base year and in the second follow-up provided baselines for the study of two important transitions experienced by the NELS:88 cohort: the transition from elementary or middle school to high school (baseline = base year) and the transition to postsecondary education or entry into the labor market (baseline = second follow-up).

2.3 Base-year through Second Follow-up Student Cognitive Test Batteries

In addition to the student questionnaire, students completed a series of achievement tests for each wave of the study at their in-school or off-campus survey sessions. The combined tests, described below, covered four subject areas and included 116 items to be completed in 85 minutes. The four subject areas included:

1. Reading Comprehension (21 questions, 21 minutes)

   This subtest contained five short reading passages or pairs of passages, with three to five questions about the content of each passage. Questions tested the students' ability to understand the meaning of words in context, identify figures of speech, interpret the author's perspective, and evaluate the passage as a whole. One version of the reading test was administered in the base year, and two versions in the first and second follow-ups.

2. Mathematics (40 questions, 30 minutes)

   Test items included word problems, graphs, equations, quantitative comparisons, and geometric figures. Some questions could be answered by simple application of skills or knowledge; others required that the student demonstrate a more advanced level of comprehension and/or problem solving. One version of the mathematics test was administered in the base year, and three versions in the first and second follow-ups.

3. Science (25 questions, 20 minutes)

   The science test contained life science, earth science, and physical science/chemistry questions and placed emphasis on the student's understanding of underlying concepts rather than on his or her retention of isolated facts.

4. Social Studies: American History/Citizenship/Geography (30 questions, 14 minutes)

   The social studies test included three categories of questions: American history, citizenship, and geography. The American history questions asked about important issues and events in political and economic history from colonial times through the recent past. Citizenship items quizzed students on the workings of the federal government and the rights and obligations of citizens. The geography questions touched on patterns of settlement and food production

shared by various societies.

The Educational Testing Service (ETS) developed the assessment batteries for all three NELS:88 in-school waves, including one test form for the base year and six forms for both the first and second follow-ups. The difficulty level of the mathematics and reading questions differed on each of the six follow-up forms, and each sample member's test form was determined by his or her scores on the base-year and/or first follow-up mathematics and reading tests. Freshened students and prior-round nonrespondents received the intermediate version of the tests.

The multilevel design of the NELS:88 achievement tests guarded against ceiling and floor effects that can occur when testing spans four years of schooling. This adaptive approach tailored the difficulty of the reading and mathematics tests to the ability of the respondent, thereby leading, given limitations in testing time, to a more accurate measurement than a single-level design. The following tables present the content and process areas for the NELS:88 cognitive tests in reading (table 2.3-A), mathematics (table 2.3-B), science (table 2.3-C), and social studies (table 2.3-D).

Table 2.3-A.—Base-year to second follow-up cognitive test specifications in reading: Content by process and test form: 1988-1992

Process	Number of items		
	Literary	Science	Social Studies/Other
Reproduction of Detail			
8th Grade	3	1	–
10th Grade Low	3	1	–
10th Grade High	2	1	1
12th Grade Low	3	1	1
12th Grade High	–	–	1
Comprehension of Thought			
8th Grade	1	1	1
10th Grade Low	1	1	1
10th Grade High	3	1	2
12th Grade Low	–	2	4
12th Grade High	–	1	8
Inferences and/or Evaluative Judgements			
8th Grade	10	1	3
10th Grade Low	10	1	3
10th Grade High	9	1	1
12th Grade Low	6	1	3
12th Grade High	4	3	3

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study, 1988–2000 (NELS:88), 1988-1992.

Table 2.3-B.—Base-year to second follow-up cognitive test specifications in math: Content by process and test form: 1988-1992

	Number of items				
Process	**Arithmetic**	**Algebra**	**Geometry**	**Data Analysis/ Probability**	**Advanced Topic (e.g., precalculus, analytic geometry)**
Skill/Knowledge					
8th Grade	10	5	1	1	–
10th Grade Low	12	4	2	–	–
10th Grade Medium	9	3	–	1	1
10th Grade High	6	3	–	2	2
12th Grade Low	10	4	2	–	–
12th Grade Medium	7	2	–	1	1
12th Grade High	1	2	–	1	2
Understanding/Comprehension					
8th Grade	6	7	3	3	–
10th Grade Low	7	6	3	2	–
10th Grade Medium	6	6	3	2	–
10th Grade High	3	7	2	3	2
12th Grade Low	6	5	3	3	–
12th Grade Medium	4	6	4	2	–
12th Grade High	1	5	7	1	3
Problem Solving					
8th Grade	3	–	–	–	1
10th Grade Low	3	–	–	–	1
10th Grade Medium	3	2	2	–	2
10th Grade High	2	2	3	–	2
12th Grade Low	4	–	2	–	1
12th Grade Medium	4	3	5	–	1
12th Grade High	2	4	9	1	1

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

Table 2.3-C.—Base-year to second follow-up study cognitive test specifications in science: Content by process and test form: 1988-1992

Process	Number of items				
	Earth Science	Chemistry	Scientific Method	Life Science	Physical Science
Skill/Knowledge					
8th Grade	5	2	–	3	–
10th Grade	3	2	–	2	1
12th Grade	3	3	–	3	1
Understanding/Comprehension					
8th Grade	2		1	2	–
10th Grade	2	2	1	2	1
12th Grade	1	1	3	1	–
		–			
Problem Solving					
8th Grade	1	3	2	2	–
10th Grade	–	3	1	3	2
12th Grade	–	3	1	2	4

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

Table 2.3-D.—Base-year to second follow-up cognitive test specifications in social studies: Content by process and test form: 1988-1992

Process	Citizenship/ Government	American History	Geography
8th Grade	13	14	3
10th Grade	12	19	3
12th Grade	8	15	3

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

Various achievement test scores, both normative and criterion-referenced, are reported in NELS:88. Available NELS:88 scores (including IRT-estimated Number Right scores, IRT theta scores, achievement quartiles, proficiency scores, and continuous probability of proficiency scores) are most fully described in appendix H (pp. H-31 – H-38) of the *NELS:88 Second Follow-Up Student Component Data File User's Manual* (Ingels, Dowd, Baldridge, Stipe, Bartot and Frankel, 1994, NCES 94-374). The psychometric basis for the scoring is described in the *Psychometric Report for the NELS:88 Base Year Through Second Follow-Up* (Rock and Pollack, 1995, NCES 95-382). The psychometric report also provides information about test reliability and validity and test specifications.

2.4 First and Second Follow-up Dropout Questionnaires

In the first follow-up, NELS:88 project staff administered a dropout questionnaire to sample members who, according to data gathered through administration of a status screener, were not in an

academic program leading to a high school diploma. This group included sample members who had received a GED or other alternative certification.

In the second follow-up, sample members who were not enrolled in a diploma-granting program and who furthermore had not obtained a GED or other alternative certification completed the dropout questionnaire (sample members with a GED or other certification completed the second follow-up student questionnaire and early graduate supplement). An interviewer was normally present at the group and individual survey sessions while students completed the hour-long, self-administered dropout questionnaire. The first follow-up questionnaire was available in English only, and the second follow-up questionnaire was available in both English and Spanish.

The dropout questionnaires collected data about the following areas:

- the last school attended by the sample member and the school's climate;
- reasons for leaving school, and actions school personnel, parents, and friends took when the respondent stopped going to school;
- the sample member's likelihood of returning to and graduating from high school; and
- the sample member's current activities, employment history, and future plans.

The research team designed the dropout questionnaire to facilitate comparisons with the NELS:88 first and second follow-up student questionnaires and the HS&B 1982 dropout questionnaire. Item overlap between the NELS:88 dropout and student questionnaires will permit NELS:88 data users to compare the school environment and experiences, family life and background, aspirations, and self-perceptions of students and dropouts. The overlap of 1982 and 1992 dropout items will facilitate comparison of contemporary dropouts with those of a decade before (see Ingels and Dowd 1995).

In both rounds, dropouts also completed the 85-minute cognitive test battery described in Section 2.3. Because of the difficulty in collecting test data from dropouts and because data from many dropouts were collected in telephone interviews that precluded testing, the NELS:88 second follow-up achieved a comparatively low (41.7 percent) weighted cognitive test completion rate for dropouts.

2.5 Supplemental Student Questionnaires

2.5.1 First and Second Follow-up New Student Supplements

For the first and second follow-ups, sample members who were first-time NELS:88 participants—due to freshening or previous ineligibility or nonparticipation—completed the new student supplement questionnaire, which was available in English and Spanish.[14] The self-administered supplement took approximately 15 minutes to complete and gathered the same basic demographic information (such as birth date, sex, family socioeconomic status, and race/ethnicity) that the base-year questionnaire had gathered for other students and their families. Because of the unchanging nature of the data, the follow-up surveys did not include questions on these topics again.

[14] In the second follow-up, survey staff also administered the new student supplement to a number of first follow-up freshened students who had completed a first follow-up student questionnaire but had not completed a new student supplement in 1990.

2.5.2 Second Follow-up Early Graduate Supplement

NELS:88 participants who graduated from high school or who obtained equivalency certification, such as the GED, before the spring 1992 data collection completed the early graduate supplement to the second follow-up student questionnaire. This supplement documents the reasons for and the circumstances of early graduation, the adjustments required to finish early, and respondents' activities compared with those of other school survey members. Instrument developers modeled the items for the NELS:88 early graduate supplement on the items used in the HS&B sophomore cohort early graduate supplement administered in the HS&B first follow-up in 1982.

2.6 Questionnaires for the Student Sample in the Out-of-School Rounds

By the time of the third follow-up in 1994, very few NELS:88 8th-grade cohort members remained in high school. This meant that while previous questionnaires (and tests) had been administered in group settings in school and optically scanned, a different mode of data collection was now required. The dominant administration form for all 1994 sample members was a one-on-one telephone interview, in a computer-assisted format (CATI, or computer-assisted telephone interviewing). The design of the 1994 questionnaire therefore departs from that of the prior rounds. By moving to an electronic format, key information could be preloaded into the interview, and automated consistency checks could be built into the interview process, minimizing the missing/inconsistent data retrieval and backend editing tasks which were an important element of the in-school rounds.

2.6.1 NELS:88 Third Follow-up Student Interview

Just as the form of the third follow-up questionnaire in 1994 differed from the form of the earlier instruments, the content differed as well, as the sample members followed diverse pathways in their transition from high school to postsecondary education or to work. Instrument developers designed the third follow-up questionnaire to focus mainly on postsecondary access and employment and to elicit valid contemporary information about these topics while maintaining as much continuity as possible with the prior NCES youth transition studies, NLS-72 and HS&B. Specific content areas included academic achievement, feelings about respondents' postsecondary institution and/or job, detailed work experience, work-related training, and family structure and environment.

Researchers field-tested the NELS:88/94 instrument in 1993 and refined it for the full-scale study based on recommendations made at the November 1993 Technical Review Panel (TRP) meeting. Members of the NELS:88 TRP included academic researchers, policy analysts, and representatives of various government agencies.

The research team conducted the NELS:88/94 interviews primarily by telephone, using CATI technology. For those cases, however, where the respondent was unable or unwilling to complete an interview over the telephone, a paper questionnaire was either self- or field-administered. The CATI system presented the questionnaire items to the interviewer on a series of screens, each with one or more questions. Between screens, the system evaluated the responses and used the results to route the interview to the next appropriate question. The system also applied a series of cross-checks to the responses, such as valid ranges, data field size and data type (e.g., numeric or text), and consistency with other answers or data from previous rounds. In addition, when the interviewer encountered problems, the system could suggest prompts to use in eliciting a better or more complete answer.

The 1994 study followed the progress of the NELS:88 cohort as sample members moved to a wide array of postsecondary activities. The study addressed issues of employment and postsecondary

access, and it sustained continuing trend comparisons with NLS-72 and HS&B. Specific content areas, described below, included family structure, high school and postsecondary academic achievement, employment experience, work-related training, environment, and locating. See the *NELS:88/94 Methodology Report* (NCES 96-174) for the CATI instrument code, which contains question text and interviewer instructions and information about preloaded data and flow. The NELS:88/94 Electronic Codebook (ECB) and the *NELS:88 Second Follow-Up: Student Component Data File User's Manual* (NCES 94-374) also contain question text for the third follow-up instrument. In addition, the facsimile of the NELS:88/94 instrument is available on the NCES NELS:88 Web site: http://nces.ed.gov/surveys/nels88/. For a summary of outcomes in 1994 covering the thematic areas listed below, see Sanderson, Dugoni, Rasinski and Taylor, 1996, *National Education Longitudinal Study 1988-1994 Descriptive Summary Report With an Essay on Access and Choice in Postsecondary Education*, (NCES 96-175).

Family structure. Family formation has been an integral component of NELS:88 since the second follow-up survey. This section of the NELS:88/94 instrument collected data on household composition, marital status, number of times married, date of first marriage, number of children, and the children's birth dates.

High school completion. Approximately 16 percent of the NELS:88 cohort had not completed high school by August of 1992. Roughly 8 percent were still enrolled and 8 percent were high school dropouts. This section collected high school information for those sample members who had not completed high school at the time of the last interview and included questions about completion status, last high school attended, dates of enrollment, highest grade attended and completed, type of high school program, type of degree/certification offered by program, and GED completion date.

Postsecondary school access and achievement. The third follow-up of NELS took place two years after most NELS sample members graduated from high school; thus, many had enrolled in a postsecondary school (e.g., community college, 4-year program). For those who had taken classes or enrolled in a postsecondary program, this section asked questions about each postsecondary institution the sample member attended, including level and control of institution, cost of tuition, dates attended, stopout activity (i.e., whether the respondent had taken time off from school before returning to the classroom), major/field of study, certificate/degree type, and certificate/degree completion and date. In addition, the questionnaire collected financial information, such as types and amount of financial aid received, and employment while enrolled.

Employment experience. The employment and income of NELS sample members, both those who were concurrently enrolled in postsecondary school and those who were not enrolled, are important to better understand the economic returns of education. This section collected information about the sample members' employment since the last interview, including spells of employment, number of jobs, job title and type of business, hours worked and income, apprenticeships, benefits, satisfaction, and expected occupation and income at age 30.

Work-related training. In addition to, or in place of formal education, workers often require specialized skills in order to do their jobs. Employers are increasingly turning to on-the-job training as a means for teaching employees new skills and competencies. The NELS questionnaire identified those who received on-the-job training and asked them about the type and amount of training, where the training took place, and how closely the training was related to their job. It also collected information about occupational licenses.

Environment. Noneconomic returns to society, such as civic involvement, are outcomes also thought to be correlated with education. The NELS:88/94 instrument included questions about the leisure

activities of sample members, such as time spent watching television, and participation in sports or religious activities. It also included items on community service and voting behavior. In addition, the questionnaire asked a set of questions about sexual activity.

Locating. The questionnaire collected locating information to aid in tracing the sample members for the next follow-up study. Items included sample member's current address and telephone number; addresses, telephone numbers, and relationship of two contacts; and driver's license information.

2.6.2 NELS:88 Fourth Follow-up Student Interview

The research team conducted the field test and full-scale NELS:88/2000 interviews both by telephone using CATI and in person using computer-assisted personal interview (CAPI) technology. In preparation for the development of the CATI/CAPI instrument, the team developed a comprehensive set of data elements from a thorough review of the data elements provided in the study solicitation, the data elements relationship to earlier administrations of NELS:88 and other elements of the education longitudinal study series, and the elements relevance to current research and policy issues. From the set of data elements, instrument developers structured the CATI/CAPI instrument by identifying section topics and determining the progression of items within sections. They then designed individual interview items with several goals in mind: (1) use prior NELS:88 items when feasible; (2) ensure consistency with prior NELS:88 items when items were not identical; and (3) identify and prepare wording for item verifications and probes, as necessary. Finally, they refined interview items for the full-scale study based on feedback from the members of the fourth follow-up study's TRP.

Despite different data collection methods, the CATI and CAPI interviews were programmed identically. The CATI/CAPI system software facilitated the preloading of full-screen data entry and editing of "matrix-type" responses. The system presented interviewers with screens of questions to ask respondents, with the software guiding the interviewer and respondent through the interview. The program skipped inapplicable questions automatically, based on prior response patterns and preloaded information. It also suggested wording for probes when a respondent provided a response that was out of range for a given item and displayed special screens or other prompts when the interviewer entered inconsistent or incomplete information. Preloaded data from the earlier administrations of NELS:88 minimized the interview burden on respondents and dictated the flow of many portions of the instrument.

The NELS:88/2000 instrument comprised 10 sections: current activities, employment, job-related training, high school completion, postsecondary education, adult education, family formation, income and expenses, other outcomes, and race-ethnicity/residence. The content of these sections is described below. For greater detail, refer to the facsimile and flow chart for the NELS:88/2000 instrument on the NCES NELS:88 Web site: http://nces.ed.gov/surveys/nels88/. For a summary of outcomes in 2000 covering the thematic areas listed below, see Ingels, Curtin, Kaufman, Alt and Chen, 2002, *Coming of Age in the 1990s: The Eighth-Grade Class of 1988 12 Years Later*. (NCES 2002–321).

Current activities. This section asked questions about the respondents' main activities at the time of the interview. These items provided the foundation for much of the remainder of the survey instrument, and the information was useful in identifying important subsets of the population. The section asked about sample members' current activity status (e.g., student, employee, homemaker, etc.) and, based on that status, collected information about unemployed sample members and current and former military service.

Employment. Capturing employment information for NELS participants who both did and did not enroll in postsecondary education is important to better understand the rate of economic return to individuals and society for various levels of education. The NELS employment items collected data on

job title, duties, salary, hours worked per week, job satisfaction, and autonomy for currently held job for pay or most recent job if not currently working.

Job-related training. Consensus grew over the past decade on skills required for the work force. The new flexible work force will require workers who have formal educational training and who are continuously learning new skills and competencies, some of which may be validated with formal state or professional licensure and certification. To ensure accurate recall periods and to more closely target specific opportunities for training, this section asked about job-related training received in the last 6 months of the current (or most recent) job. Interviewers questioned members of the sample cohort who received such training, on the structure, purpose, and impact of their job-related training activities.

High school completion. A key milestone in a young person's life is completion of high school. By 1994, more than 87 percent of the NELS:88 cohort had earned a high school diploma or GED. The NELS:88/2000 interview updated high school completion information for those who had not completed high school by 1994 or who were not interviewed in 1994. Interviewers asked students who had obtained a GED their reasons for completing their high programs with the equivalency exam and whether they participated in a GED study program.

Postsecondary education. The postsecondary data items in the fourth follow-up of NELS, conducted 8 years after most NELS participants graduated from high school, provide important information for addressing issues of student access to postsecondary education, patterns of persistence within the system, and postsecondary educational attainment. This section collected the names, locations, and IPEDS codes[15] for all postsecondary institutions attended by sample members since high school graduation, degrees or certificates obtained, date of degree/certificate, and field of study. This section also collected information about postsecondary education experiences and aspirations.

Adult education. This section explored the ways in which respondents engage in learning beyond formal postsecondary education and job-related training. Young adults have a wide range of educational opportunities at their disposal from a variety of sources, and they engage in them for a variety of reasons. For example, sample members may take classes over the Internet, participate in continuing education courses at local schools and museums, and even obtain private tutors. In fact, creating lifelong learners is one of the important objectives of elementary and secondary education.

Family formation. The fourth follow-up of NELS is a rich resource of information regarding historical trends in family formation that are directly comparable to the HS&B and NLS-72 cohorts. This section collected data on current marital status, including the dates of marriage and how marriages ended (if applicable); household composition; number of dependents and children; and birth dates of the oldest and youngest children.

Income and expenses. Considering the substantial earnings advantages of education, economic returns are one of the most important outcomes of education. This interview collected information about respondents' and their spouse's or partner's income in 1999, 1998, and 1997. This section also collected other measures of financial condition, such as current housing status and public assistance.

Other outcomes. This section collected information about community integration and healthy behaviors—factors that are commonly believed to be correlated with education and labor market

[15] NCES' Integrated Postsecondary Education Data System (IPEDS) surveys all primary providers of postsecondary education in the U.S. on enrollment, faculty, staff, and finances. Each postsecondary institution is assigned a unique unit identification number. In NELS:88, IPEDS codes are available only on the restricted use files.

outcomes. Questions focused on integration with and involvement in the community (e.g., volunteerism, voting behavior); questions about health-related issues included cigarette and alcohol use.

Race-ethnicity/residence. While prior NELS:88 instruments asked for respondents' racial-ethnic status, the fourth follow-up collected multiracial responses and included greater specificity for Asian and Native Hawaiian/Pacific Islander races, in accordance with new federal standards for the collection of information on race and ethnicity. This section also included an item on the racial-ethnic diversity of the respondents' work and residential communities comparable to HS&B. The section concluded with information on the respondents' current place of residence, which can be used, in conjunction with locations during the previous follow-up and base-year surveys, to examine the mobility of young adults.

In addition to the CATI/CAPI interview just described, instrument developers created an abbreviated instrument expressly to conduct difficult-to-complete interviews with sample members. They developed this interview in two content-identical modes—hard copy and electronic versions—to collect data from sample members who either could not complete interviews by telephone (e.g., sample members without telephones or who were incarcerated) or would not complete telephone interviews (e.g., refusals). The abbreviated instrument focused on respondents' current activities, postsecondary education, and work experiences.

2.7 Base-Year through Second Follow-up School Administrator Questionnaires

The primary purpose of the school administrator questionnaire was to gather general descriptive information about the educational setting and environment associated with the individual students selected for participation in NELS:88. This school information describes the overall academic climate in terms of specific school practices and policies, as well as enrollments and educational offerings. The information obtained through the school administrator questionnaire provides supplemental data to the student questionnaire so that student outcomes can be considered in terms of school measures. The NELS:88 base-year school survey provided a national probability sample of 1988 8th-grade schools and a stand-alone school data set. *Because the first and second follow-up school samples do not constitute a national probability sample of schools, the first follow-up and second follow-up school administrator data should be used only as contextual data for student-level analyses.* While it is not correct to generalize 1990 and 1992 NELS:88 school administrator data to all the nation's high schools, NELS:88 does supply nationally representative samples of 1990 sophomores and 1992 seniors (as well as of 1988 8th-graders two and four years later). Student-administrator matches may therefore be used, as long as the student remains the unit of analysis.

In each survey wave, the NELS:88 school principal, headmaster, or other knowledgeable school official designated by the school administrator completed the self-administered school administrator questionnaire (which was 40 minutes in length in the base year, 60 minutes in the first follow-up, and 45 minutes in the second follow-up). For the first follow-up, the research team also designed an abbreviated version of the questionnaire for telephone administration to nonresponding principals. The base-year through second follow-up questionnaires contained similar content. Topics covered included:

- General school characteristics, such as grade span, school, and 12th grade enrollment sizes, and school control and demographic characteristics.
- General student characteristics for the modal grade of the survey cohort, including average daily attendance rates, ethnic and racial composition, percentage of students with limited English proficiency, and numbers of students receiving special school services.

- Teaching staff characteristics encompassing such areas as the number of full-time and part-time faculty, departmentalization of faculty, salary levels, and evaluation of teachers.
- School policies and programs, including requirements for minimum competency and proficiency tests, and programs for language minority students.
- School governance and climate, such as administration practices, school reforms, types of parental involvement, student behavioral problems in school, and areas of principal's control.

The research team designed the school administrator questionnaire so that the first several sections could be answered either by the school principal or by a designee who was able to provide the requested information. Only the principal could answer the last section, which asked for his or her subjective opinions regarding the school environment.

2.8 Base-Year through Second Follow-up Teacher Questionnaires

The NELS:88 teacher component was designed to provide teacher information that can be used to analyze the behaviors and outcomes of the student sample, including the effects of teaching on longitudinal student outcomes. ***The design of this component does not provide stand-alone analysis samples of teachers, but instead provides contextual data for analyses at the student level***. The teacher component supports comparison of specific teacher characteristics and practices to the learning context and educational outcomes of sampled students. The component also supplies teacher ratings or evaluations of student sample members. The teacher questionnaire is the critical instrument for investigating the student's specific learning environment. At the same time, a limitation of the teacher component design is that even within a given subject, such as mathematics or science, there are gaps in coverage of some of the period during which learning is taking place (e.g., 9^{th}- and 11^{th}-grade teachers were not surveyed).

In both the base year and first follow-up, selected teachers completed a 45-minute, self-administered questionnaire. The teachers selected were responsible for instructing sampled students in two of the four cognitive test subjects: mathematics, science, English, and social studies (American history, civics [citizenship/government], and geography). (The four two-subject combinations normally selected for students were either mathematics or science combined with either English or social studies). In the first follow-up, when possible, NELS:88 project staff chose teachers who taught the sample member in one of the same two cognitive test areas that were chosen for that student in the base year. In some cases, however, students were not enrolled in classes in the same subject areas as they were during the base year; NELS:88 project staff therefore chose a teacher from another one of the four subjects to evaluate them. In the second follow-up, if the student was enrolled in either a mathematics or science class, survey staff again selected a teacher for one of the two subjects to respond to a 30-minute questionnaire. In all three survey waves, interviewers asked teachers to respond to the questionnaire items in relation to a specific list of sampled students enrolled in their classes.

The teacher questionnaire sought to illuminate questions of the quality, equality, and diversity of educational opportunity by obtaining information in the following four content areas:

- Teacher's assessment of the student's school-related behavior and academic performance, educational and career plans and goals. Respondents completed this section with respect to the sample members they instructed in a particular subject.

- Information about the class the teacher taught to the sample member (e.g., track assignments, instructional methods, homework assignments, and curricular contents). This section of the instrument included classroom topic coverage items ("opportunity to learn" items) that articulate with the cognitive tests.
- Information about the school social climate and organizational culture (e.g., teacher autonomy, participation in determining school policy, and relationships with the principal).
- 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).

A validation study of NELS:88 teacher reports on instructional content, instructional strategy, and goals was conducted in the second follow-up (Burstein et al. 1995). Teachers completed daily logs over a 5-week period, describing their instructional practices, and the research team obtained copies of teachers' textbooks and other artifacts, such as homework, quizzes, classroom exercises, projects, and exams, which they then coded. The team compared this information with survey responses.

The authors found that teachers reported curricular topics more accurately for upper-level than for lower-level courses and that survey data "reveal reasonably accurately whether a topic has been taught not at all, for only a few periods, for a week or two, or for several weeks." They found that survey data "present an accurate picture of the instructional strategies used most often by teachers, and they provide some indication of how teachers combine strategies during instruction." The authors' analysis suggests that instructional goals, however, "cannot be validly measured through national surveys of teachers."

2.9 Base-Year and Second Follow-up Parent Questionnaires

Instrument developers designed the self-administered parent questionnaire to collect information from parents about factors that influence educational attainment and participation. The objective of the parent questionnaire was to provide data that could be used primarily in the analysis of student behaviors and outcomes; it was designed only secondarily as a data set of parents. The questions focused on family background, socioeconomic characteristics, and the character of the home educational support system. In addition, the parent instrument collected data related to parental behaviors and circumstances with which the student may not have been familiar, such as parental education and occupation. It also contained more sensitive questions about income, postsecondary educational costs and financial aid decisions, and religious affiliation. In both the base year and the second follow-up, the parent questionnaire instructed the parent or guardian who was most knowledgeable about the sample member's educational activities and related behaviors to complete the questionnaire. Accordingly, the parent respondent was self-selected.

The parent questionnaire covered the following thematic areas:

- Information about the family's background (base year and second follow-up). In this section of the questionnaire, respondents identified their relationship with the student or dropout sample member, provided data on the family size and composition, and answered questions about their employment situation and occupation, race, and language background and skills.
- Information about the teenager's school life (base year and second follow-up). This section elicited parental knowledge of key characteristics of the teenager's educational situation and collected data on the forms of interaction between the school and parent.

- The teenager's family life (base year and second follow-up). This section of the questionnaire asked parents about the decision-making process within the household and the kinds of interaction between the respondent and teenager. It included several sensitive questions about community life and drug and alcohol use by the teenager.
- Opinions about the teenager's school (base year only).
- The teenager's postsecondary plans (second follow-up only). This section covered parental aspirations for the teenager, preparations for postsecondary education, and plans for the teenager's transition to the work force.
- The teenager's plans for the future (second follow-up only). This section covered parental educational aspirations for the teenager.
- Financial information and educational costs. This section included items about family income and financial preparations for the teenager's postsecondary education.
- Supplemental questions for parents new to NELS:88 in the second follow-up (second follow-up only). The research team administered the final section of the second follow-up parent questionnaire only to parents who had not participated in the base-year parent survey either because the parent or guardian was a base-year nonrespondent or because the student was added to the sample in the first or second follow-up. This section included a number of questions asked in the base-year parent survey for which new data were not required from base-year respondents. These items covered family characteristics, size, and composition in 1988, parent education, and parent age.

In the base year, a small number of parents were interviewed by telephone. In the second follow-up, a greater proportion of parents completed telephone interviews. In both surveys, the research team took a number of steps to minimize mode effects, including training interviewers to adapt questionnaire items so that they were intelligible when read over the telephone and asking parents to read along in the questionnaire during the interview if they had a copy of the self-administered questionnaire.

2.10 Transcript Studies

2.10.1 Second Follow-up Transcript Component

In the second follow-up, the research team collected high school transcripts for members of the contextual sample (students for whom contextual school and teacher data were collected), all eligible sample members who were dropouts (including GED recipients) or early graduates, and sample members who were in the 12th grade in 1992 and ineligible for all three waves of NELS:88. Collecting the high school transcripts facilitated two important research efforts:

- the validation of certain data—including high school course taking, course grades, and attendance data provided by sample members in their responses to the first follow-up and second follow-up questionnaires; and
- the investigation of course-taking patterns by sample member characteristics, and the relationship of such patterns to sample members' postsecondary activities and achievement.

The research team also conducted the transcript study to enable comparisons with the transcripts studies from HS&B (1982) and NAEP (1987, 1990, 1994, 1998, and 2000) (see Ingels and Taylor 1995,

for notes on using the various transcript data sets for trend analysis). In reviewing the transcripts, team members abstracted the following data elements:

- Student-level items, including number of absences per year, rank in class and class size, date student left school, reason student left school (graduated, transferred, etc.), cumulative GPA, and standardized scores for the PSAT, SAT, ACT, College Board Achievement tests, and Advanced Placement tests.
- Course-level items (for courses taken in grades 9 through 12), including course title, department, and number; year, grade level, and term course taken; number of credits earned; and grade awarded.

2.10.2 Fourth Follow-up Postsecondary Education Transcript Study

The research team conducted a postsecondary education transcript study after the fourth follow-up of NELS:88 in 2000, to add richness and depth to the academic data collected during the third and fourth follow-up studies. The study primarily sought to gather data on course-taking behavior and postsecondary achievement. All fourth follow-up respondents who reported any postsecondary education were included in the transcript collection study.

The research team collected data on

- institutional characteristics of institutions attended by sample members (name, location, level, control, description),
- degrees obtained (degree earned, field of study of degree), and
- course-taking behavior.

2.11 Sources of Further Information on NELS:88 Instrumentation

2.11.1 Questionnaires

English-language questionnaires for the base year through third follow-up were reprinted in the various base-year through second follow-up user's manuals. The 1990 NELS:88 Spanish language questionnaires appear in appendix K of the *NELS:88 First Follow-Up Final Technical Report* (Ingels, Scott, Rock, Pollack and Rasinski 1994; NCES 94-632; available on the NCES Web site). The 1992 Spanish language questionnaires were reprinted in Ingels, Scott, and Taylor (1998) the *NELS:88 Base Year Through Second Follow-Up Final Methodology* Report (available in the NCES Working Paper Series, [NCES 98-06]; also, the document can be downloaded from the NCES Web site).

In addition, English-language questionnaires appear on the CD-ROM release of the NELS:88/2000 data. Questionnaires can also be viewed on the NELS:88 Web pages on the NCES Web site (http://nces.ed.gov/surveys/nels88).

For an evaluation of the performance of the 1988-92 questionnaires—specifically, comparison of student and parent reports, comparison of student and teacher responses, and comparison of responses across the in-school survey waves—readers are referred to:

McLaughlin, D.H., and Cohen, J. (1997). *NELS:88 Survey Item Evaluation Report* (NCES 97-052). Washington, DC: National Center for Education Statistics.

2.11.2 Cognitive Test Battery

Although the student data file user's manuals contain some information on the cognitive test battery, the most comprehensive account of the NELS:88 achievement tests can be found in:

Rock, D.A., and Pollack, J.M. (1995). *Psychometric Report for the NELS:88 Base Year Through Second Follow-Up* (NCES 95-382). Washington, DC: National Center for Education Statistics.

The 1995 report covers the base year (1988), first follow-up (1990), and second follow-up (1992); however, further detail on the base-year tests can be found in:

Rock, D.A., and Pollack, J.M. (1991). *Psychometric Report for the NELS:88 Base Year Test Battery* (NCES 91-468). Washington, DC: National Center for Education Statistics.

Description and documentation of the psychometric properties of the first follow-up tests can be found in:

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M., and Rasinski, K. (1994). *NELS:88 First Follow-Up Final Technical Report* (NCES 94-632). Washington, DC: National Center for Education Statistics. (See chapter VI.)

2.11.3 Transcript Studies

More detailed information on both the high school transcript summary variables on the NELS:88 public-use files and the restricted-use transcript component data file can be found in:

Ingels, S.J., Dowd, K., Taylor, J.R., Bartot, V., Frankel, M.R., and Pulliam, P.A. (1995). *NELS:88 Second Follow-Up: Transcript Component Data File User's Manual* (NCES 95-377). Washington, DC: National Center for Education Statistics.

Related transcript sources that may be of interest include:

Alt, M.N., and Bradby, D. (1999). *Procedures Guide for Transcript Studies* (NCES Working Paper Series 1999-05). Washington, DC: National Center for Education Statistics. (Available on-line at the NCES Web Site or in hard copy from NCES.)

Ingels, S.J., and Taylor, J.R. (1995). *National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data* (NCES Working Paper Series 1995-06). Washington DC: National Center for Education Statistics. (Available on-line at the NCES Web site, from ERIC, or in hard copy from NCES.)

Documentation for the NELS:88 Postsecondary Education Transcript Study will be released in the near future.

2.11.4 Other works cited in the chapter

Burstein, L., McDonnell, L.M., Van Winkle, J., Ormseth, T., Mirocha, J., and Guiton, G. (1995). *Validating National Curriculum Indicators*. Santa Monica: RAND.

Ingels, S.J., and Dowd, K.L. (1995). *Conducting Trend Analyses: HS&B and NELS:88 Sophomore Cohort Dropouts* (NCES Working Paper Series, No. 95-07). Washington, DC: National Center for Education Statistics. (Available from NCES Web site, from NCES in hard copy, or from ERIC.)

Chapter III
Sample Design, Weighting, and Design Effects

This chapter addresses three broad topics: the National Education Longitudinal Study of 1988 (NELS:88) sample design, weighting, and variance estimation and design effects. More specifically, this chapter describes the design and procedures used for selecting schools and students for the NELS:88 base-year data collection; details subsampling and other sample selection activities in subsequent waves of data collection from the first through the fourth follow-up studies; and describes the sample weighting, and reports on standard errors and design effects. Although the 1988-2000 public and restricted-use ECBs (N0P and N0R) contain only the fourth follow-up weights, this chapter fully documents the weights produced in all prior NELS:88 rounds. Full documentation of past weighting has been provided both to consolidate this information in one place and to supply a context for understanding the continuities and differences between weights for the 2000 round and the prior NELS:88 rounds. This chapter does not provide information about the sample design or weighting for the 1990-92 NELS:88 High School Effectiveness Study (HSES). However, sources of HSES information are listed at the end of the chapter. The chapter also does not discuss sampling or weighting for the NELS:88 postsecondary transcript study. This information will be contained in a separate document available with the restricted-use data from that special study.

3.1 Objectives of the NELS:88 Sample Design

The following section outlines the objectives of the NELS:88 sample design, from its base-year inception through the fourth follow-up. Beginning as a straightforward, two-stage stratified sample, the NELS:88 sample design grew in complexity with each subsequent wave of the survey.

The sample design for the *base year* in 1988 was similar in many respects to the designs used in both the National Longitudinal Study of the High School Class of 1972 (NLS-72) and High School and Beyond (HS&B), the two previous studies in the National Center for Education Statistics' (NCES') education longitudinal study series. The principal difference between NELS:88 and the two previous NCES studies is that, in its base year, NELS:88 sampled a cohort of 8th-graders rather than high school students. Also included in the NELS:88 sample was a supplementary sample of Hispanic and Asian/Pacific Islander students (and their parents and teachers) sponsored by the Office of Bilingual Education and Minority Languages Affairs (OBEMLA, currently called the Office of English Language Acquisition), U.S. Department of Education. From a national frame of about 39,000 schools with 8th grades, a total of 1,734 schools were selected, of which 1,057 ultimately participated[16] (the realized sample). The realized sample of students selected from these schools was 24,599.

Most students changed schools after 8th grade. Conducted two years after the 8th-grade base year, the NELS:88 *first follow-up study* was designed to

[16]However, owing to loss of data in transit and other problems, usable student data were obtained for only 1,052 schools.

- Sample approximately 21,500 students who were in the 8th-grade sample in 1988 (including nonrespondents to the base-year data collection). This longitudinal cohort was to be distributed across 1,500 first follow-up schools containing 10th-grade students.
- Constitute a valid probability sample of all students enrolled in the 10th grade in the 1989-90 school year. This entailed "freshening" the sample with students who were 10th-graders in 1990 but who were not in the 8th grade during the 1987-88 school year.
- Retain members of the base-year cohort who had dropped out of school (sample with certainty) to maximize the number of cases available to study this policy-relevant group.
- Retain nonrespondents from the base-year data collection to minimize nonresponse bias.
- Include a sample of excluded students, or in other words, those who were deemed ineligible for base-year data collection because of physical, mental, or linguistic barriers that prevented them from participating. (Eligibility status for this group was reassessed; some students' eligibility status, particularly those with limited English language proficiency, changed over time. Newly eligible students were added to the first follow-up student sample, and demographic and school enrollment information was obtained for them. The entire group [those who became eligible over time, and those who remained ineligible] became part of an "expanded sample" and was used to supply a bias correction factor for key estimates such as 8th-grade cohort dropout rates.)

The NELS:88 *second follow-up study* was designed to

- Constitute a valid probability sample of all students enrolled in the 12th grade in the 1991-92 school year. This entailed freshening the sample with students who were 12th-graders in spring term of 1992 but who were not in the 8th grade in the United States in the 1987-88 school year. Additionally, it was necessary to reassess the eligibility status of selected students classified as ineligible in previous waves and to include them in the second follow-up cohort if they were now deemed eligible.
- Like the previous follow-up study, retain dropouts and nonrespondents to minimize nonresponse bias and maximize the number of cases in the dropout analysis group.
- Retain the maximum number of Hispanics, Asian/Pacific Islanders, and American Indian/Alaska Native sample members from the first follow-up sample for policy analysis purposes.
- Provide a contextual data sample that would be distributed across no more than 1,500 schools from which school administrator, teacher, and transcript data would be collected.

It was hoped that these goals could be achieved with minimal loss to both sample efficiency and effective sample size.

To control costs in the NELS:88 *third follow-up study* (NELS:88/94), subsampling was instituted to reduce the second follow-up sample of 21,635 to some 15,964 sample members. In order to ensure a sufficient numbers of cases for analysis, rarer subgroups were retained at a higher rate. In the NELS:88 *fourth follow-up* (NELS:88/2000), two instances of further subsampling took place, including the undersampling of third follow-up study nonrespondents.

Three public-use ECBs are currently available that correspond to the NELS:88 student samples in the following way: one ECB (N2P) is limited to the 1988-92 in-school rounds and reflects all students

who participated at any time in the 1988-92 waves of NELS:88. A second public-use ECB (N4P) contains 1988-94 data, based on 1994 sample participants only. A third public-use ECB (N0P) contains 1988-2000 data, based on 2000 wave sample participants only. This final ECB includes 12,144 study respondents, including 93 third follow-up nonrespondents. (Corresponding to each of the three public-use ECBs is a restricted-use ECB: N2R, N4R, and N0R). A fourth restricted-use ECB will contain postsecondary education transcript data and weights for sample members.

3.2 Sample Design for the 8th Grade Panel

The NELS:88 base-year survey employed a two-stage, stratified sample design, with schools as the first-stage unit and students within schools as the second-stage unit. Within each stratum, schools were selected with probabilities proportional to their estimated 8th grade enrollment to achieve virtual self-weighting. In addition, schools were oversampled in certain special strata to ensure that policy-relevant subgroups would be adequately represented in the sample. Within each school, approximately 26 students were randomly selected.[17] In schools with fewer than 24 8th-graders, all eligible students were selected. Because of the incidence of small schools in the NELS:88 sample, the average within-school sample size for the base year was 25 students, of which 23 students ultimately participated. The number of students sampled in each school ranged from 1 to 73 students. From a national frame of about 39,000 schools containing the 8th grade, a target sample size of 1,032 schools was selected. Some 1,052 schools—815 public and 237 private—participated and provided usable 8th-grade student data.

Because of the greater representation of small private schools and the impact of a within-school strategy of oversampling Hispanics and Asian/Pacific Islanders, there is considerably greater variability in within-school sample size in the NELS:88 base year than in the HS&B base-year sample. The target population for the NELS:88 base year consisted of all public and private schools containing the 8th grade in the 50 states and District of Columbia. Excluded from the NELS:88 sample were U.S. Department of the Interior-supported Bureau of Indian Affairs (BIA) schools, special education schools for persons with severe disabilities, area vocational schools that did not enroll students directly, and U.S. Department of Defense dependents schools. In order to minimize burden on individual participating schools, schools selected for the U.S. Department of Education's National Assessment of Educational Progress (NAEP) were excluded from the NELS:88 sample. The student population excluded students with severe mental disabilities, students whose command of the English language was not sufficient to understand the survey materials, and students with physical or emotional disabilities that presented difficulties to survey participation.

3.2.1 Sampling Frame

When designing a sampling frame, either an explicit or an implicit list of elements to be sampled can be used. For NELS:88, the creation of an explicit list of all 8th-grade students in the United States would have been an impossible task. However, a comprehensive list of 8th-grade students is implied as a function of a comprehensive list of all schools with 8th-grades. Project staff, with the assistance of participating schools, developed an implicit list of students from public and private schools in the United States. It was important that the list of schools be complete and accurate, especially with variables used in subsequent sample stratification.

[17] Typically, these students included 24 regularly sampled students and two students from the OBEMLA supplementary sample (Asian/Hispanic oversample). Because some schools had small 8th-grade student counts with less than 24 students, as well as student transfers, the average sample size across the 1,052 schools was 25.1.

Investigation of various sources indicated that the most readily accessible source for a complete and accurate frame was the database compiled by Quality Education Data, Inc. (QED), a marketing research firm located in Denver, CO. The QED database contained public and private schools, including both parochial and nonparochial private schools. It also contained Census information about whether a school's location was urban, suburban, or rural, and this information was used to stratify the schools. The QED list did not contain school-level information about the racial-ethnic composition of public or private schools. Instead, racial-ethnic composition data for use in constructing the NELS:88 public school sampling frame was obtained from Westat, Inc.[18]

Westat obtained Black and Hispanic percentages directly from district personnel in public districts that, according to the QED list, had large proportions of Black or Hispanic students. These data were compiled only for public schools in the primary sampling units of the 1986 National Assessment of Educational Progress (NAEP). In all, less than one-half of the 8th-graders in the NELS:88 sampling frame came from schools for which such racial composition data were available. However, these partial data allowed the creation of sampling strata containing public schools with large percentages of Black or Hispanic students. In addition, data from the QED list allowed identification of schools as public, Catholic (private), or other private for stratification purposes. The stratification procedures are discussed in more detail in the following sections.

3.2.2 Stratification

The sampling frame was sorted in such a way as to create groups of schools, called strata, that were contiguous on the frame. Each stratum contained schools that were relatively similar in terms of certain variables deemed relevant to the survey's objectives. The actual selection of schools occurred independently within each stratum. Schools were stratified by superstrata and substrata. First, schools were sorted into combinations of school type and geographic region (superstrata). Next, substrata were formed according to values on an urbanization variable (i.e., whether the school was in an urban, suburban, or rural area) and according to the minority classification discussed above. Minority substrata were not created for private schools.

Schools within substrata were sorted in order of their estimated 8th-grade enrollment size. The sort order alternated from ascending to descending from one substratum to the next. Table 3.2.2 indicates the number of schools in the sampling frame for each stratum. Note that some schools were classified as ineligible after they were sampled and contacted. These schools were excluded from the sample, and the tabled values do not reflect these schools (see Section 3.3.1 for a discussion of excluded schools). However, subsequent descriptions of the sample do account for the ineligible schools. Therefore, the number of schools reported in some tables varies slightly from the numbers reported here.

The divisions that formed the public superstrata were equivalent to the regions used by the Census Bureau. Single states that formed superstrata were excluded from the divisions.[19] The regions that formed the private school strata were the same as the Census regions, except that one state was excluded from the Northeast region.

[18] As part of NAEP, Westat obtained data from the U.S. Office for Civil Rights (OCR) and other sources that identified schools with a combined Black and Hispanic enrollment of greater than 19 percent.

[19] For example, New York formed its own superstratum and was removed from the Northeast sample division.

Table 3.2.2.—Numbers of schools in NELS:88 base-year sampling frame and number of schools sampled, by sampling strata: 1988

School Type	Schools in Frame	Schools Sampled
Total public and private	38,866	1,734
Public Schools		
Total Public	22,818	1,350
Northeast/Middle Atlantic	3,650	273
East North Central	4,101	224
West North Central	3,217	100
South Atlantic	2,604	225
East South Central	1,976	91
West South Central	2,994	168
Mountain	1,629	76
Pacific	2,647	193
Private Schools		
Total Private	16,048	384
Catholic, Suburban/Rural		
Northeast	1,233	33
North Central	1,762	32
South	539	10
West	521	9
Catholic, Urban		
Northeast	515	17
North Central	1,450	28
South	569	11
West	362	6
Other Private		
Northeast	1,072	69
North Central	3,038	52
South	2,808	71
West	2,179	46

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

3.2.3 Allocation of Numbers of Schools to be Sampled

The number of public schools to be selected for the core sample from each superstratum was set to be proportional to the aggregate estimated 8th-grade enrollment of all the schools in that superstratum. For this calculation, the 8th-grade enrollment in each school was estimated by dividing the enrollment figure from the QED list by the number of grades in the school; this procedure implicitly assumes an

equal number of students in each grade in the school. The allocation of the sample size to substrata within the public school superstrata was proportional to the sum of a measure of size (MOS) of the schools in the substrata. The MOS was proportional to the 8th-grade enrollments. (The calculation of the MOS is discussed in Section 3.2.4.)

The determination of the numbers of schools to be selected from each of the private strata reflected a compromise between competing analytic needs. Private schools as a whole were oversampled relative to public schools. In addition, an augmentation of private schools in one northeastern state was included in the sample. Policy analysts are particularly interested in certain types of private schools, and oversampling these types has the obvious benefit of increasing the number of cases available for analysis, but at the cost of decreased precision for statistics based on other types of private schools. The allocation was designed to give policy analysts the minimum number of schools necessary for their work while preserving as much as possible an allocation proportional to 8th-grade enrollment. This would ensure a reasonable level of precision for estimates based on all types of private schools.

3.2.4 Selection of Schools within Strata

A sample design objective was that each student sampled from the selected schools would have an equal chance of selection. To accomplish this, an MOS was calculated for each school that was not selected by the NAEP:

$$MOS = F \times G \times \max \{24, \text{G8 enrollment}\}$$

Schools selected by the NAEP had MOS set to zero. The MOS was equal to an adjustment factor, *F*, times another factor, *G*, times the maximum of 24 (which was the desired number of regular students per school to be sampled) or the estimated 8th-grade enrollment of the school. The factor *F* varied from school to school and was designed to adjust for the fact that the NAEP did not select schools with equal probability. *F* was set equal to the reciprocal of *1-P*, where *P* was set to equal each school's probability of selection into the NAEP.[20] This ensured that choosing schools with probabilities proportional to the MOS would yield school selection probabilities proportional to estimated 8th-grade enrollments. The latter is desirable because if the school selection probabilities are proportional to 8th-grade enrollments and if 24 (or all students if fewer than 24 are enrolled) are selected at random from each selected school, then all students have equal probabilities of selection.

The factor *G* is present in the format above to ensure that a sufficient number of other-private school students are included in the sample. Many of the other-private schools had estimated 8th-grade enrollments considerably under 24, and if the factor *G* was not present, then the number of sampled students in other-private schools would be lower than acceptable. The factor *G* was set to 1 for all schools in all strata except for the superstratum other-private. For schools in the latter superstratum, *G* was set to 1 if the estimated 8th-grade enrollment was 8 or more, and *G* was set to 0.5 if the estimated 8th-grade enrollment was less than 8. The effect of *G* is to undersample small private schools where very few

[20] For each school,
P(NELS) = probability of selection into NELS
P(NELS/NAEP) = probability of selection into NELS given selection into NAEP
P(NELS/not NAEP) = probability of selection into NELS given nonselection into NAEP
P = probability of selection into NAEP
Also, let ENROLL denote an estimate of the number of students in the 8th grade in the school. Then,
$P(NELS) = P(NELS/NAEP) \times P + P(NELS/\text{not NAEP}) \times (1-P)$
Note that P(NELS/NAEP) = 0
Thus, $P(NELS) = P(NELS/\text{not NAEP}) \times (1-P)$
If P(NELS) is set proportional to ENROLL, then P(NELS/not NAEP) is proportional to ENROLL/(1-P).

students could be sampled. With a fixed school sample size, this has the effect of increasing the number of large other-private schools, thus increasing the total number of other-private students in the sample.

The selection of the public schools was accomplished using systematic sampling with random starts in each public superstratum and with the sampling intervals in each superstratum proportional to the MOS. The selection of the private schools was accomplished using systematic sampling with random starts in each private substratum and with the sampling intervals proportional to the MOS. Use of systematic sampling produced the beneficial effect of implicit stratification by 8th-grade enrollment within each substratum.

3.2.5 Design Allowance for School Nonresponse

Despite the best efforts of data collection personnel, not all schools selected for the survey agreed to participate in the study. To avoid potential problems related to this nonresponse,[21] project staff drew extra schools in the initial selection process. These extra schools were brought into the sample as necessary to maintain desired sample sizes despite nonparticipation in some schools. The extra schools were chosen at random from the same superstratum and substratum as nonresponding schools. The sample drawn was larger than the sample initially intended; schools were randomly assigned to two pools, with Pool 1 containing the target sample and Pool 2 containing backup schools. The research team attempted to obtain cooperation from Pool 1 selections. However, when cooperation was not possible, an additional school was randomly selected from Pool 2 with the same superstratum and substratum as the nonresponding school. This procedure had the effect of controlling the number of cooperating schools from each superstratum and substratum.

Schools selected randomly within each substratum were alternately assigned to pools, with each school having an equal chance of being in Pool 1 or Pool 2. All schools from Pool 1 were selected for the study. Where the number of responding schools in a stratum from the first Pool was below a prespecified target number, second pool schools were selected.[22] It is important to note that not all Pool 2 schools were fielded. Once the target number of schools within a stratum was obtained, additional Pool 2 schools were not considered further. School weights were derived based on the number of Pool 1 and Pool 2 schools that were contacted, without respect to the pool to which the school was initially assigned.

The final sample size consisted of all Pool 1 schools and all Pool 2 schools from which cooperation was requested; Pool 2 schools that were not contacted were not counted. The final sample size (adjusted for numbers of ineligible schools) was used as the denominator of the unweighted response rate for schools. The sample design weight for each extra (Pool 2) school that was brought into the survey was calculated in the same manner as the weights for the Pool 1 schools (i.e., as the reciprocal of the selection probability conditional on the final sample size for the school's superstratum and substratum).

3.2.6 Selection of Students

The basic sampling procedure resulted in the selection of up to 24 students per school, or all of the 8th-grade students in the school if they numbered fewer than 24. An additional procedure was implemented to augment this basic sample of 24 students per school with an oversample of Asian/Pacific Islander and Hispanic students. The target was to achieve a total oversample of 2,200 additional students with these racial-ethnic characteristics.

[21] The potential problems associated with nonresponse include systematic errors in statistics calculated from data collected from participating schools, and decreases in the size of the sample from which data are obtained.

[22] See Section 3.3.2 for a discussion of the specified target numbers (see also *Table 3.3.2*).

The student sampling procedure can be described as follows: first, three lists of 8th-graders were obtained from each participating school, one of Asian/Pacific Islander students, one of Hispanic students, and one of all other students. Second, random samples of Asians, Hispanics, and others were independently selected from each of the three lists. Sample sizes were calculated using the following formulae:

$$nH = (CS \times CH \times NH/F) + (24 \times NH/N),$$

$$nA = (CS \times CA \times NA/F) + (24 \times NA/N),$$

$$nO = 24 \times NO/N,$$

where *nH, nA, and nO* are sample sizes for the Hispanic, Asian, and other students. *NH, NA,* and *NO* denote the number of students on the lists of Hispanics, Asians, and others, respectively, and *N* denotes the total number of students on all of the lists. *F* denotes the first-stage selection probability of the school, *CA* and *CH* are constants used for the selection of Asian and Hispanic students, and *CS* is a constant used for the selection of Asian and Hispanic students in stratum *S*. *CA, CH,* and *CS* were constants of proportionality constructed to obtain desired total sample sizes for Asian, Hispanic, and Other students across schools.

Upper limits on *nH* and *nA* were set to ensure that the number of students per school was not larger than practical from the standpoint of the logistics of survey administration. The specifications of *CS, CA,* and *CH* were empirically determined to ensure that two goals were achieved: (1) sufficient numbers of Asian and Hispanic students were sampled, and (2) selection probabilities did not vary excessively across students. Design effects were also kept from becoming too large.

3.2.7 Sample Updating

A representative from each school submitted a list of eligible students from which a sample was drawn. These lists, called school rosters, were submitted and an initial sample was drawn, starting in November 1987. To adjust the student sampling frame for student attrition and change in the 8th-grade population of the sampled school, a sample update was conducted seven to ten days prior to the school's scheduled survey session. Field staff reviewed the sampling list on site with the school coordinator to ascertain whether all sampled students were still eligible and to ensure that transfer-ins (any student who joined the school's 8th-grade class between the time of original sampling and the time of the update) were added to a supplementary roster from which additional students would be selected. The supplementary roster was annotated for eligibility and ethnicity, and the transfer-in students were sequentially numbered. Selections for inclusion in the sample were based on the same set of computer-generated random numbers used to select the original sample and Asian/Hispanic oversamples for that particular school. While in the HS&B, base-year substitutions were made for students who were ineligible or who had died, there were no student-level substitutions in NELS:88.

3.2.8 Selection of Contextual Data Samples

In its base year, NELS:88 surveyed not only students, but also school administrators, teachers, and parents. In the first follow-up, students, dropouts, school administrators, and teachers were surveyed. In the second follow-up, students, dropouts, school administrators, teachers, and parents were surveyed. In addition, two new contextual components were added in 1992: a high school transcript survey and a survey of high school course offerings and enrollments. This section describes the selection of the base-year school administrator, teacher, and parent survey samples. Section 3.5.4 describes contextual survey samples selected after the base year.

Selection of School Administrators. The head administrators (principals, headmasters, and headmistresses) of all eligible 8th-grade schools in the universe of schools constituted the universe of school administrators. A head administrator from each school in the NELS:88 sample was asked to complete a questionnaire.

Selection of Teachers. All full- and part-time teachers who were teaching classes in mathematics, science, English/language arts, and social studies to eligible 8th-graders in the winter/spring of 1988 were included in the NELS:88 universe of 8th-grade teachers. The actual sample was restricted to teachers who provided instruction in the four subject areas to the selected sample of 8th-grade students within the sampled schools. Thus, there was no need to construct a formal universe list of 8th-grade mathematics, science, English, and social studies teachers prior to their selection. In cases where the teacher had changed between the autumn and spring terms, the spring-term teacher was designated as the preferred respondent. To achieve the objective of linking information from the teacher to data about individual students in the NELS:88 sample, two teachers were selected to respond to the teacher questionnaire for each student.

Respondents were selected to complete the teacher questionnaire for each student based on the assignment of two curriculum areas per school included in the NELS:88 base-year sample. Specifically, each of the sample schools was assigned one of the following combinations of curriculum areas:

- Science and English,
- Science and social studies,
- Mathematics and English, or
- Mathematics and social studies.

Each sampled student's current teacher in each of the two designated curriculum areas was selected to receive a teacher questionnaire.

The assignment procedure was designed to achieve approximately balanced representation of the four combinations of curriculum areas across the sampling variables of school type and levels of urbanicity and/or minority population. Additionally, there was an attempt to balance assignments within geographical categories and by school size. Finally, the assignment process was intended to ensure representation of mathematics or science, and English or social studies teachers in all base-year sampled schools.

Once the data file listing all sampled schools was compiled, it was sorted in the order of sample selection; that is, by geographical category within school type, then by urbanicity/minority level, by whether the school was selected initially as a sample school or a replacement school, and finally by an MOS. Next, the four subject area combinations were randomly ordered. The ordering obtained by randomization was (1) mathematics and social studies, (2) mathematics and English, (3) science and English, and (4) science and social studies. The ordered schools were assigned to repeating cycles of the above order of subject combinations.

Following the assignment of curriculum combination areas to sampled schools and the selection of the student sample in a participating school, a matrix of student-subject-teacher information was obtained from school records. For each student-curriculum combination (subject), the following information was collected:

- Class identification (e.g., usually period number or hour),
- Course title, and
- Name of the student's current teacher in that subject.

In completing the teacher matrix, the school coordinator was asked to report the current teacher, or where there were multiple current teachers, to report the teacher who had the greatest assigned responsibility for teaching the sampled student (if equal responsibility, a teacher was randomly picked). The assignment of subject-matter pairs to schools ensured that data were collected from two teachers of each student (assuming that there were more than two teachers for the 8th-grade class and that both the student's teachers chose to participate in the study) and that survey response burden for teachers in the school was limited.

Because of the universality of the four subject matters in the required curriculum of the 8th grade, all sampled students were enrolled in classes in the assigned subject combination during some portion of the 1987-88 school year. Thus, no subject substitution was necessary. However, occasionally, a student was enrolled in more than one spring-term class in a particular subject. When this was so, the following decision rule was invoked to determine which class would be entered on the teacher matrix:

1. When there are two or more candidate classes in the same subject for a given student, take the course in which the student will have spent the most class time between the start of school and the survey day; if this rule is not sufficient to eliminate all but one of the candidate classes, select the class that involves the most advanced subject matter.

2. In other cases, where more than a single teacher could be selected for a student, the teacher with the greatest assigned responsibility was chosen to complete the teacher questionnaire.

The use of this sampling scheme for the NELS:88 base year resulted in the identification of varying numbers of teacher-respondents per participating school, ranging from 1 to 19, with an average number of 5.5 per school. It should be noted that the resulting NELS:88 base-year sample of teacher-respondents did not constitute a statistical or representative sample of 8th-grade teachers for analysis and reporting purposes. Rather, the results of this questionnaire were intended to provide information about student-related characteristics, teacher practices, and curriculum exposure, which may affect longitudinal student outcomes. Once data collection had been completed, the sample was further restricted to teachers of base-year participants; that is, data collected from teachers of base-year nonparticipants were systematically excluded from the data files.

Selection of Parents (or Guardians). Conceptually, the universe of parents of 8th-grade students consisted of all parents or legal guardians of 8th-grade students in the winter/spring of 1988. The selection of parents or guardians thus did not require the construction of a formal universe or list. One parent questionnaire was sought per student, regardless of whether the student resided in a one- or two-parent home (or split time between parents with joint custody, in the case of divorced parents). Once the student sample was selected, the parent or guardian who was "best-informed" about the child's educational activities was asked to complete a NELS:88 parent questionnaire. Thus, the parent respondent was essentially self-selected.

No effort was made to identify parents who had more than one chance of selection (that is, had more than one child in the 8th grade). After parent and student data had been collected, the parent sample was further restricted to the parents/guardians of participating base-year students. Thus, parent data from the base-year nonparticipants were systematically excluded from the final data file.

3.3 Ineligibility and Exclusion and School Nonresponse Bias

3.3.1 Exclusions from the Sample

Exclusion of Students. Students who were judged by their schools to be incapable of meaningful participation in NELS:88—students with severe disabilities, or so lacking in proficiency in the English language that they could not be validly assessed in English—were deemed ineligible. Of the potential sample of spring 1988 8th-graders, 5.3 percent were excluded from the sample by virtue of such ineligibility. To better understand how excluding persons with mental handicaps, insufficient English-language proficiency, and severe physical and emotional problems affects population inferences, data were obtained on the numbers of students excluded from the base-year sample and on the reason for exclusion. In addition, as will be described in Section 3.4.2, a special study of these students was done in the first and second follow-ups, when their eligibility status was re-assessed, and when appropriate, base-year excluded students were included in the follow-up samples.

Students who were educated at home or in private tutorial settings and those who had dropped out of school before reaching the 8th grade also fall outside the NELS:88 base-year sample. The size of the pre-eighth grade dropout population in the winter/spring of 1988 is uncertain. NCES reported that 12 percent of dropouts aged 16 to 24 had completed six or fewer years of school (Frase 1989). However, more than 31 percent of Hispanic dropouts aged 16 to 24 had completed only six or fewer years of school. This finding both confirms the fact that there is a sizable group of students who leave school before entering 8th grade and suggests that the biasing effect of this phenomenon on NELS:88 data may be more pronounced for some subgroups than for others.

Exclusion of Schools. Just as certain students were considered to be ineligible for the base-year sample, so too were certain kinds of schools. The eligible populations of schools were restricted to "regular" schools in the United States, private as well as public. Excluded from the sample were Bureau of Indian Affairs (BIA), U.S. Department of the Interior, schools; special education schools for students with disabilities; area vocational schools that did not enroll students directly; and schools for dependents of U.S. military personnel overseas. Additionally, a sample list school was considered ineligible if the school no longer existed (closed or merged) or did not enroll any 8th-grade students in the spring term of 1988. Most of the sample list schools declared ineligible were schools that had closed, or small, private schools that had no 8th-grade students enrolled in the spring 1988 term. Finally, a school was declared ineligible if it was established after the final sampling frame was constructed.[23]

These exclusions did not have a large impact on estimates made from the base-year sample. Information from the Department of Education's Common Core of Data and other sources suggest that about 90 percent of American Indian and Alaska Native children attend schools not affiliated with BIA.[24] Investigators should take this degree of undercoverage into account when attempting population estimates. If this group is substantially different from American Indian 8th-graders not attending BIA schools, a substantial bias in estimates may result. (Additional information on the characteristics of American Indian/Alaska Native schools, staff, and students may be found in Pavel and Curtin 1997; Pavel, Curtin, Thorne, Christenson, and Rudes 1995).

[23] The sample frame represented information current through April 1987.

[24] In this situation, "affiliated" means school either operated directly by the BIA, U.S. Department of the Interior, or operated under BIA contract to tribal organizations or governments.

Similarly, fewer than 10,000 8th-graders attended Department of Defense Dependents' Schools (DODDS) serving dependents of U.S. personnel overseas in the 1987-88 school year. This estimate suggests that fewer than 0.3 percent of all 8th-grade students were in DODDS schools.

3.3.2 School Nonresponse

Sample realization in the NELS:88 base year is summarized below in table 3.3.2. As in HS&B, approximately 70 percent of initial schools selected agreed to participate.

Table 3.3.2.—NELS:88 base-year school sample selection and realization: 1988

	Target N	Pool 1				Pool 2	Pool 1 and 2
		Eligible	Ineligible	Agreed		Replacements	Participating Schools
				n	%		
Total	1,032	1,002	30	698	69.7	359	1,057*
Public	800	774	26	522	67.4	295	817
Catholic	95	91	4	70	76.9	34	104
Private	137	137	0	106	77.4	30	136

*1,057 schools participated at some level, though owing to loss in transit, usable student data were received for only 1,052. For 1,035 schools, both student and school administrator data were received.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

School-level nonresponse is a serious concern because it carries over into successive rounds of NELS:88. Students attending schools that did not cooperate in the base year were not sampled and had little or no chance of selection into the follow-up samples. To the extent that students at noncooperating schools differ from students at cooperating schools, the student-level bias introduced by base-year school noncooperation persists during subsequent waves. Nonresponse adjustments to weights are an attempt to compensate for bias in the estimate for a particular subgroup; they do not adjust for nonresponse bias within subgroups.

In the base year, nonresponding schools were asked to supply information about key school questionnaire variables, and virtually all did so. Based on these data, analysis of school-level nonresponse suggests that, to the extent that schools can be characterized by size, control, organizational structure, student composition, and other characteristics, the impact of nonresponding schools on school-level estimates is small.[25] Readers interested in more information about the analyses of school nonresponse rates and bias for the NELS:88 base year should refer to the *NELS:88 Base-Year Sample Design Report* (Spencer et al. 1990). School nonresponse was not assessed in the first or second follow-ups for two reasons. First, there was practically no school-level nonresponse; institutional cooperation levels approached 99 percent in both rounds. Second, the first and second follow-up samples were

[25] The use of school questionnaire variables to assess bias in estimates concerning characteristics of the student population is not entirely straightforward. Still, to the extent that school characteristics are closely related to the characteristics of the students attending them, estimates based on school questionnaire data can serve as reasonable proxies for more direct estimates of student-level unit nonresponse bias.

student-driven, unlike the two-stage initial sample design in the base year. Hence, even if a school refused in either the first or second follow-ups, the individual student was pursued outside of school.

3.4 Changes to the 8th Grade Panel Sample Design after the Base Year

Several changes to the NELS:88 sampling design were implemented after the base year. These changes included

- Subsampling the 8th grade cohort and freshened sophomore samples in 1990;
- Conducting studies of excluded students in 1990 and 1992;
- Freshening the 1990 sophomore and 1992 senior samples; and
- Subsampling schools in which contextual components were administered in 1992.

This section discusses the subsampling of the 8th-grade cohort and freshened sophomore samples in 1990, the studies of excluded students in 1990 and 1992, and the subsampling of schools in which contextual components were administered in 1992. Section 3.5 discusses the freshening of the 1990 sophomore and 1992 senior samples.

The sample design for the 1990 wave of data collection attempted to maximize statistical efficiency while limiting cost. The general sample design strategy for the longitudinal cohort component of the first follow-up sample involved subsampling students selected for the base year with nonzero probabilities related to characteristics of their 1990 schools. Base-year students who had dropped out of school between 1988 and 1990 were subsampled with certainty (that is, their probabilities of selection were set to 1). Base-year students attending school in 1990 were subsampled with probabilities related to the number of other base-year students attending the same school. Base-year students who were reported to be attending a school with at least 10 other base-year students were sampled with certainty. All other students were sampled with probabilities greater than zero, but less than 1.

Including nonrespondents, the NELS:88 base-year sample comprised 26,432 students. Of these, 96 were deemed out of scope for the 1990 first follow-up (including students who had died or moved out of the United States). Among the remaining 26,336 students, 348 were found to have dropped out of school; all of these students were selected into the first follow-up with certainty (probability of selection equal to 1).[26]

The remaining pool of 25,988 students was distributed among 3,967 schools.[27] As had been anticipated, the distribution of these students among schools was highly skewed. It was found that approximately 75 percent of the students (19,568 of 25,988) were attending approximately 23 percent (908 of 3,967) of the schools; each of these schools included at least 11 base-year students. All of these 19,568 students were included in the first follow-up subsample with certainty. The remaining 6,420 students were distributed among 3,059 schools with 10 or fewer members of the base-year sample. Their

[26] The 348 dropouts comprise 250 dropouts whose status was confirmed by the student's home, 58 sample members whom the school reported to have dropped out but field interviewers could not locate, and 40 students who were institutionalized. The latter group are not necessarily dropouts in the strict sense of the first follow-up dropout definition because in some cases they were receiving academic instruction. However, they were grouped with the dropouts to ensure that they would remain in the first follow-up sample with certainty.

[27] When the school a student was attending could not be identified, a separate "school" of size one (i.e., one person) was created. This was the case for 221 students who could not be located and ten students who were in home study. Hence, the number of actual schools was 3,736.

sampling probabilities for the first follow-up depended on the number of base-year students the school contained. The efficiency of this design relative to one with no subsampling at all was 66.5 percent.[28]

3.4.1 Subsampling the 8th Grade Cohort and Freshened Sophomore Samples

After the initial selection of the longitudinal cohort, the combined longitudinal-freshened sample was further subsampled.[29] The students dropped from the first follow-up as a result of subsampling were also excluded from the second follow-up. Two categories of sample members were subsampled: (1) students who had transferred out of the school from which they had initially been selected for the first follow-up sample; and (2) first follow-up nonrespondents who were classified as potential dropouts.

Transfer students were subsampled as a cost-saving measure. Because of the large number of transfer students and the high costs of obtaining questionnaires from them, NORC selected a 20 percent subsample of transfer students in the spring of 1990. Of the 1,991 transfers, 386 were retained and 1,605 were dropped from the sample.

A 50 percent subsample of "potential dropouts" was drawn after the end of the regular data collection period in the spring of 1990. The subsampling encompassed those students who had not been located in the data collection phase and those who had been absent at the time of in-school data collection session(s). Those selected into the subsample were the object of renewed follow-up efforts to identify any "hidden dropouts" in these categories of cases. There were 742 "potential dropout" cases, of which 357 were retained in the sample and pursued in the final data collection period of the study. In the course of final data collection, substantial numbers of these "potential dropouts" (75 of the 357 subsample members) were confirmed as having been dropouts at the time of their school's survey session and were included as part of the first follow-up dropout study; the remaining 282 were identified as still in school.

As a result of this subsampling, the longitudinal cohort and the 10th-grade freshened student samples were reduced by 1,997 cases, yielding a first follow-up sample size of 20,706 (table 3.4.1).[30] While this number represents the number of sample members included on the public-release data file, additional students—the 340 members of the sample of base-year ineligibles found to be eligible or out of scope in the first follow-up were added to the second follow-up's re-release of the first follow-up sample files (see the following paragraphs for a description of the sample of base-year ineligibles).

3.4.2 1990 Study of Excluded Students

The NELS:88 base-year sample excluded students for whom the NELS:88 survey instruments would be unsuitable (i.e., students with a mental disability and students who were not proficient in English) and students whose physical or emotional problems would have made participation in the survey

[28] The measure of efficiency was computed as $1/(1+RV) \times 100\%$, where RV is the relative variance of the weights required to compensate for the different rates of subsampling.

[29] The process referred to here as "freshening" added students who were not in the base-year sampling frame, either because they were not in the country or because they were not in 8th grade in the spring term of 1988. The 1990 freshening process provided a representative sample of students enrolled in 10th grade in the spring of 1990. The 1992 freshening process provided a representative sample of students enrolled in 12th grade in the spring of 1992. Section 2.5 of this report describes the freshening process.

[30] The provisional first follow-up sample size of 20,706 was amended to include 340 base-year ineligible students who were reclassified as eligible or out of scope in the first follow-up. Additionally, data for 23 sampling errors found among the students freshened into the sample or out of scope in the first follow-up, as well as four additional sampling errors, have been deleted. Finally, 179 first follow-up freshened dropouts have been excluded from the public-use files. Accordingly, the revised first follow-up sample size is 20,840.

Table 3.4.1.—NELS:88 first follow-up study sample, by race breakdown[1]: 1990

	First Follow-up Initial Selections	Freshened Sample	Dropped-in Final Subsampling[2]	Final Sample
All	21,474	1,229	1,997	20,706[3]
Asian/Pacific Islanders	1,367	89	141	1,315
Hispanics	2,828	246	323	2,751
American Indians	278	28	32	274
Blacks	2,265	235	280	2,220
White	14,349	554	1,061	13,842
Missing/Refused	387	77	160	304

[1] Figures represent the first follow-up constructed variable frequencies. This variable—race identified at the time of sampling—is not the same variable included on the data files and reported in the ECBs. This variable was used because it was the only race variable that was constructed for initial sample members dropped in final subsampling.

[2] 1,821 members of the 8th-grade longitudinal cohort and 169 freshened 10th-graders were dropped in phase 3 subsampling. In addition, 7 members of the 8th-grade longitudinal cohort were discarded because they were selected in error during the base year.

[3] The final sample here is based on the original (1992-93) release of the 1990 first follow-up student file. The sample size was revised in the second follow-up study to account for sampling errors and the inclusion of base year ineligible students.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1990.

unduly difficult. Data were obtained on the numbers of such ineligibles to facilitate inferences to the larger population that includes such persons. About 5.3 percent of the students at base-year sample schools were excluded from participation. Of these, 57 percent were excluded because of mental disability, another 35 percent because of language barriers, and 8 percent because of physical disability.

There were several reasons for adding a sample of ineligibles to the first follow-up design. One such consideration was a change in eligibility rules between the base year and first follow-up. Because a Spanish translation of the first follow-up questionnaire was developed and because the requirement that standardized tests be administered was waived for those who could not complete them in English, it was feasible for some of the base-year ineligibles to take part in the first follow-up who could not have taken part in the base year. Another consideration was the need to accommodate eligibility change, as another means of providing for a probability sample of 1992 12th-graders.[31] Students whose ineligibility status had changed between 1988 and 1990 also could be surveyed in the first follow-up. However, even for those excluded base-year students who still could not complete the NELS:88 instruments, additional demographic information was collected in order to better describe any undercoverage biases, and school

[31] While, in general, the tendency is for certain classes of ineligible students to become eligible (for example, speakers of other languages come to be proficient in English), in rare instances eligible 1987-88 8th-graders had become ineligible in the first or second follow-ups (for example, because of mental or physical problems engendered by an accident). NORC treated students who were outside the United States in the 1991-92 school year as out of scope for the second follow-up, but they retained their overall sample eligibility.

enrollment status information was collected to provide a more accurate estimation of a national dropout rate between grades 8 and 10.

Because the ineligibles had been excluded prior to the base-year sample selection, NORC simulated the selection of a base-year sample that included these ineligibles. Within each base-year sample school, the same within-school sampling rates that had been used in selecting the base-year sample students were applied. A total of 674 ineligibles were selected for the simulated base-year sample by the following procedure, with a final sample size of 653: the eligibility status of these students was reassessed, their school enrollment status and basic demographic characteristics were determined, and student questionnaire data were obtained from those deemed able to complete a questionnaire. These data have been released with the rest of the first follow-up sample in the final release of the second follow-up data on the 1994 ECB. Student questionnaire data from those who were successfully surveyed are included in the combined base-year/first follow-up/second follow-up data release.

3.4.3 1992 Study of Excluded Students

In the second follow-up, base-year ineligibles who were found to be eligible in the first follow-up—whether dropouts or students—were treated as full cohort members. The base-year ineligibles who were found to be still ineligible in the first follow-up constituted the bulk of the sample in the 1992 study of excluded students. Table 3.4.3 provides summary statistics for this special study. Two additional groups of students were also included in this component. First, a small number of first follow-up students selected for freshening were declared ineligible and were therefore included in the study of excluded students. Second, a small number of sample members who were eligible for participation in the base year became ineligible for the first follow-up or the second follow-up. These sample members were a generally rare group to whom mentally or physically incapacitating events occurred, rendering them ineligible for the second follow-up main study but now eligible for the study of ineligibles.

Table 3.4.3.—Summary of final 1992 statuses for excluded students from the NELS:88 base-year study: 1988-1990
[in unweighted percent]

Reason for 1988 exclusion	Eligible		Ineligible		Not determined		Total
	n	%	n	%	n	%	
Total	334	57.2	186	31.8	64	11.0	584
Language barrier	125	71.0	22	12.5	29	16.5	176
Physical disability	13	56.5	9	39.1	1	4.3	23
Mental disability	166	50.3	140	42.4	24	7.3	330
Unknown reason	30	54.5	15	27.3	10	18.2	55

NOTE: Excludes cases sampled in error and those out of scope (dead or out of country) for the 1992 round.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1990.

The second follow-up study of excluded students pursued essentially the same objectives as the first follow-up base-year ineligible study. Because the competence of any of these previously excluded students may change between waves, their eligibility status was reassessed through informed sources (typically, a special education teacher, guidance counselor, or English-as-a-Second-Language teacher). Additionally, complete school enrollment status information was obtained, as well as confirmation of basic demographic characteristics.

3.4.4 Expanded Sample

The studies of excluded students allow for some deviation from the traditional definition of survey participation and provide the opportunity to calculate dropout rates that account for survey ineligibility. The HS&B and NELS:88 base-year definition of survey participation was, at minimum, completion of the student questionnaire. Nonrespondents, or those for whom there was no completed questionnaire in a round, received no final (nonresponse-adjusted) weight and do not appear in the final data file, except for summary demographics and status flags.

The alternative approach is to acknowledge a second level of participation in the study, based on whether school enrollment status information and the most basic sociodemographic classification variables can be obtained. To generate school retention and dropout statistics that account for those incapable of participation in the strict sense of questionnaire and test completion (and those who are capable but did not participate), basic sociodemographic and school persistence information was collected through school personnel or by proxy (usually a parent or guardian) for both nonparticipants and ineligibles. A special weight was created to reflect this expanded definition of the "participating" population.

The NELS:88 second follow-up expanded sample file—which includes basic sociodemographic, school persistence, and survey eligibility information, along with cohort flags and statistical weights—allows licensed researchers to estimate dropout rates for grades 8 to 10, 10 to 12, and 8 to 12. In addition, the expanded sample files assist researchers in the exploration of undercoverage bias. Uses of the expanded sample data are found in *Dropout Rates in the United States, 1992* (McMillen et al. 1993), and *Sample Exclusion in NELS:88: Characteristics of Base-Year Ineligible Students; Changes in Eligibility Status After Four Years* (Ingels 1996). It is important to note that, because of confidentiality requirements, the expanded sample file is available for use only to licensed researchers. Analysts are also reminded that only the expanded sample weight should be used with these data. A full description of the expanded sample weight is provided in Section 3.8.2.2 of this report.

3.4.5 Longitudinal Cohort in 1992

When second follow-up tracing of cohort members was completed, it was found that the first follow-up sample (that is, the sum of base-year respondents and nonrespondents retained after first follow-up subsampling, and first follow-up freshened students) was much more widely dispersed than had been anticipated. Taking into account non-deceased in-country locatable cases, after eliminating the locations of the "known" dropouts[32] (*N*=1,564) from consideration (dropouts were sampled with certainty), the remaining eligible sample of students (*N*=18,726) was dispersed among 3,224 schools/locations.[33]

[32] In the second follow-up, dropouts were defined differently for sampling purposes than for data collection purposes. (See the *NELS:88 Second Follow-Up: Dropout Component Data File User's Manual*, Section 4.3.1, for further details regarding the definition of dropouts for data collection and questionnaire assignment.) For sampling purposes, dropouts comprised all individuals who were classified in the first follow-up as ever having dropped out—that is, dropouts (individuals who were not enrolled in school in the spring term of 1990) and stopouts (spring-term 1990 students with a recorded 1988-90 dropout episode), regardless of their school enrollment status as of the second follow-up spring-term 1991 tracing effort. Dropouts who returned to school and stopouts who remained in school were still counted as dropouts for sampling purposes, along with institutionalized individuals and dropouts identified during second follow-up tracing. For sampling, some dropouts who were out of school after tracing returned to school and were interviewed as spring-term 1992 students.

[33] Including dropouts, there were 4,788 locations. Once nonschool locations associated with dropouts, early graduates, institutionalized sample members, home-study students, and unlocatables were subtracted from the total,

It was clear that, even if no attempts were made to satisfy the goal of retention with near certainty of Hispanics, Asians, and American Indians from the first follow-up sample, the goal of achieving a cluster of students in 1,500 schools could not be met without significant losses in sample efficiency, effective sample size, or both. Table 3.4.5 shows the distribution of students eligible for second follow-up sampling (excluding dropouts) by school size, as well as the number of schools with at least one sample member who was either Hispanic, Asian, or American Indian. To achieve disproportionate retention of minority students, most of the schools containing these students would have to be selected, leaving few additional sample selections to distribute among the remaining school sites and contradicting the initial sampling plan to include with certainty any school with at least five NELS:88 sample members enrolled at the school. After consideration of several alternative allocations—taking into account the negative effects of subsampling on sample efficiency, the strong desire to retain as many Blacks, Hispanics, Asians, and American Indians as possible, and the substantial investment made in two prior rounds in obtaining student, parent, teacher, and school data for those students who would have been subsampled out—it was decided to include all first follow-up sample members in the second follow-up sample.

3.4.6 The 1992 Contextual Components Subsample

A total of 2,258 schools were identified in the second follow-up tracing of the NELS:88 first follow-up sample; 1,500 of these were targeted for contextual data collection. All 1,030 schools identified as having four or more first follow-up sample members enrolled were included in the school-level sample with certainty (i.e., probability of 1.0). Schools with three or fewer students were subjected to sampling according to the following process: a random sample of 321 of the 1,008 (probability = 0.31845) schools identified as containing one first follow-up sample member was selected for retention in the sample. A random sample of 104 of the 160 (probability=0.65) schools containing two first follow-up sample members was selected for retention. Finally, a random sample of 45 of the 60 (probability=0.75) schools containing three sample members was selected.

The school administrator, teacher, and transcript components[34] were limited to the contextual subsample of 1,500 schools. For students in the 1,500 schools selected, the full range of data—student, school administrator, teacher, parent, and transcript data—was collected; for the students in a school not among those selected, only student and parent data were collected.

Users should note that school-level data from this sample of schools, to be used in analysis with second follow-up student data, must be adjusted with a weight calculated separately for these students. If that weight is not applied, there will be a potential for systematic bias with respect to those factors associated with attendance at schools with fewer NELS:88 students. For example, students who are more likely to transfer to different schools will be underrepresented if the weight is not applied. Further details on second follow-up weighting can be found in Section 3.8.3.2.

there were 2,258 school sites. Of these, 1,008 had a cluster of one student, 160 had a cluster size of two, 60 had a cluster size of three, and 1,030 had a cluster size of four or more students.

[34] Transcripts were also collected for (1) all dropouts, dropouts in alternative programs, and early graduates, regardless of school affiliation; and (2) 1988 8th-graders who were ineligible for the base-year, first follow-up, and second follow-up surveys because of a mental or physical disability or language barrier.

Table 3.4.5.—Clustering of NELS:88 first follow-up sample members eligible for second follow-up (schools [*N*=2,258] and nonschool locations): 1990-1992

School Size	Total Number	With Asian/Pacific Islanders, Hispanics, and Indians	Without Asian/Pacific Islanders, Hispanics, and Indians
Total	3,224	1,383	1,841
1	1,974	579	1,395
2	160	70	90
3	60	25	35
4	53	35	18
5	38	14	24
6	26	17	9
7	27	17	10
8	33	20	13
9	21	10	11
10	36	22	14
11	43	31	12
12	35	20	15
13	47	37	10
14	51	35	16
15	57	41	16
16	53	37	16
17	82	48	34
18	72	48	24
19	77	58	19
20	65	43	22
21	55	43	12
22	40	31	9
23	32	27	5
24	22	21	1
25	13	12	1
26	6	6	0
27	6	5	1
28	5	3	2
29	7	6	1
30	4	2	2
31	5	5	0
32	2	1	1
33	1	1	0
34	1	1	0
35	2	2	0
36	3	3	0
37	1	1	0
38	1	0	1
40	1	1	0
41	2	1	1
44	1	0	1
45	1	1	0
50	1	1	0
53	1	1	0
60	1	1	0

NOTE: Known dropout and stopouts are not included in the numbers above.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1990.

3.5 Cross-Sectional Sample Design: 1990 and 1992

3.5.1 Freshened Sophomore Sample

The second sampling objective in the first follow-up was to create a valid probability sample of students enrolled in 10th grade in the 1989-90 school year; this goal was achieved by a process called freshening.

The freshening procedure was carried out in four steps:

1. For each school that contained at least one base-year 10th-grade student who was selected for an interview in 1990, a complete alphabetical roster of all 10th-grade students was obtained.

2. For each base-year sample member, the next student on the list was examined. If the base-year student was the last one listed on the roster, the first student on the roster was examined.

3. If the student who was examined was enrolled in the 8th grade in the United States in 1988, then the freshening process terminated. If the designated student was not enrolled in the 8th grade in the United States in 1988, then that student was selected into the freshened sample.

4. Whenever a student was added to the freshened sample in Step 3, the next student on the roster was examined and Step 3 was repeated. The sequence of Steps 3 and 4 was repeated (adding more students to the freshened sample) until a student who was in the 8th grade in the United States in 1988 was reached on the roster.

The freshening process could yield zero, one, or two or more new sample members in a given school. Altogether, 1,229 new students were added to the 10th -grade sample—on average, just less than one student per school. Some of these freshened students were dropped in the subsampling process (described above) either because they were not included in the subsample or because the base-year student to whom they were linked was not included. Some 1,043 students selected through the freshening procedure remained in the final first follow-up sample.

3.5.2 Freshened Senior Sample

The sample freshening process was once again employed in the second follow-up to ensure that 1992 12th graders who had no opportunity for selection in the base year were included, thus eliminating one of two obstacles to the cohort being a valid probability sample of 1991-92 high school seniors. (The second obstacle was the prior exclusion of some 1988 8th-graders, which is addressed in the next section.) The procedure was implemented in four steps as described above, with the exception that second follow-up freshening was also performed for students who were added to the NELS:88 cohort through freshening in the first follow-up; in other words, a first follow-up freshened student was treated like any cohort member and could bring in another student through freshening in the second follow-up.

This freshening procedure is an essentially unbiased method for producing a probability sample of students who were enrolled in the 12th grade in 1992 but who were not enrolled in the 8th grade in the United States in 1988. There is a very small bias introduced by the omission of eligible 12th graders

attending schools that included *no* students who were 8th-graders in 1988.[35] There is an additional small bias introduced by not freshening the members of the sample of base-year ineligibles. All other 1992 12th graders who qualified for the freshening sample had some chance of selection. Because each 1988 8th grader added through first follow-up freshening had a calculable, nonzero probability of selection into the base-year sample, one can calculate the selection probabilities for all students eligible for the freshening sample. Thus, the freshening procedure produces a sample that meets the criteria for a probability sample.

Implementation of student sample freshening in the first and second follow-ups was subject to a set of eligibility rules that were patterned after but not identical to those of the base year. While again students with overwhelming physical, mental, or linguistic barriers to participation were excluded, students not sufficiently proficient in English to complete the tests or regular questionnaire but able to complete the student questionnaire in Spanish were classified as eligible and asked to complete the translated instrument. (Through the first follow-up base-year ineligibles study and second follow-up study of excluded students, this broadened eligibility criterion was also applied to excluded 1987-88 8th-graders at two points in time.) Of the 366 students initially sampled through the freshening process, 288 were found to be eligible and were brought into the cohort; 266 of the 288 were identified as being eligible to participate in the second follow-up. Some 22 of the 266 (8.3 percent) were later determined to be ineligible; 8 were excluded owing to physical or mental disabilities, 13 because they had moved out of the country, and 1 for language reasons.

It also should be noted that the school sample from which school contextual data (teacher questionnaires, school administrator questionnaires, and transcripts) were collected is not identical to the school sample as used for freshening. Freshening took place at all schools at which there were NELS:88 sample members *as of the first day of the 1991-92 school year*.[36] The school sample, for purposes of collecting contextual data, comprised the 1,387 schools that (1) represented selected clusters at which NELS:88 sample members were still present in the 1991-92 school year, and (2) provided at least one completed student questionnaire.

3.5.3 Contextual Components in 1990

One important element of the contextual component sampling plan in the first follow-up was that no parent data were gathered. Chief administrators of all schools with first follow-up sample members still in attendance were, however, asked to complete a self-administered questionnaire. No effort was made to select a nationally representative school sample in 1990.

Up to two teachers of each first follow-up student sample member were asked to complete a questionnaire. To maximize longitudinal comparability of teacher data, NELS:88 first follow-up teachers for each student were selected in the same subject combinations as those selected in the base year. (In other words, if, for example a student's base year link was to an English teacher and a math teacher, if possible, that student was again, in the first follow-up, linked to the same subject combination, English and math). Freshened students were assigned the subject combination of their base-year "linked"

[35] For purposes of implementation of the freshening process, a "school" was defined as an institution whose primary purpose is the provision of instruction and that grants diplomas or certificates. This definition categorically excludes certain types of places of instruction (e.g., prison schools).

[36] Only those freshened sample members who remained in school through the spring term became members of the HS&B-comparable NELS:88 sophomore cohort. However, autumn sophomores who had dropped out by spring were surveyed in both the first and second follow-ups. While these "freshened dropouts" were included on the original first follow-up public release, in subsequent re-releases these cases appear only on the restricted-use files.

partner. If a student was enrolled in only one of the four subject areas, then only one teacher report was collected for the student.

3.5.4 Contextual Components in 1992

As a result of the dispersion of students to a large number of schools between 1990 and 1992, it was necessary to select a sample of 1,500 schools from the set of all schools attended by second follow-up students in which a full complement of school data collection activities (school administrator, teacher, transcript surveys) would take place. Of these 1,500 schools, the full complement of contextual component activities occurred in 1,374. For students attending schools other than those 1,374, only student and parent questionnaires were administered.

Chief administrators of the 1,500 subsampled schools were again asked to complete a questionnaire. Again, in 1992, no effort was made to select a nationally representative school sample.

In a departure from previous waves, only one second follow-up teacher (either mathematics or science) of each student in the contextual (1,500) school sample was asked to complete a teacher questionnaire. If a student was not enrolled in either a mathematics or science class, no teacher questionnaire was administered.

The second follow-up parent sample consisted of a subsample of the "most knowledgeable" parents or guardians of second follow-up student and dropout participants. Using parent data with the parent weight provides analysts with only a national cross-section of the self-defined parent or guardian of, among other populations, students from the 8^{th}-or 10^{th}-grade cohorts.

The subsample of parents was selected to ensure that completion rates for key subgroup populations were comparable with completion rates of other subgroups. The probability of a parent's selection for the subsample was determined by the priority of the subpopulation of the parent's teen. Subsampling occurred when 70 percent of data collection was completed, at which point all parents for whom a questionnaire had not yet been collected received a selection probability. Table 3.5.4 lists the subpopulations subsampled and their probabilities of selection.

High school transcripts were collected in 1992 for (1) the contextual sample—students attending the 1,500 sampled schools in the spring of 1992; (2) all dropouts, dropouts in alternative programs, and early graduates, regardless of school affiliation; and (3) triple ineligibles enrolled in the 12^{th} grade in the spring of 1992, regardless of school affiliation. Triple ineligibles are 1988 8^{th}-graders who were ineligible for the base-year, first follow-up, and second follow-up surveys as a result of a mental or physical disability or language barrier. NELS:88 transcript data are intended to enable licensed analysts to examine trends when contrasted to HS&B (1982) high school transcripts, and the 1987, 1990, 1994, 1998 and 2000 NAEP transcript studies.

3.6 Sample Design and Implementation: NELS:88 Third Follow-up Study

The sample for NELS:88 third follow-up was created by dividing the NELS:88 second follow-up sample into 18 groups based on their response history, dropout status, eligibility status, school sector type, race, test scores, SES, and freshened status. Each sampling group was assigned an overall selection

Table 3.5.4.—Subsampling rates for the NELS:88 second follow-up study's parent sample: 1992

Parents of Student/Dropout Cohort Subgroup	Probability of Selection
Dropouts	1.0
Private school students	1.0
White students, public schools, low socioeconomic status (SES)	1.0
Black students	1.0
Hispanic students, public schools	0.5
Asian students, public schools	0.8
Other students, public schools	0.5
White students, public schools, high SES	0.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1992.

probability. Cases within a group were selected such that the overall group probability was met, but the probability of selection within the group was proportional to each sample member's second follow-up design weight. Assigning selection probabilities proportional to the second follow-up design weight reduced the variability of the NELS:88 third follow-up raw weights and consequently increased the efficiency of the resulting sample from 40.1 percent to 44.0 percent. Table 3.6-A defines the 18 sample categories for the third follow-up study. Table 3.6-B lists the selection probabilities and follow-up sample weighting distributions for these groups. While some sample members qualified for more than one of the sample groups, each member was assigned to only one group. The groups were created in order of priority, so that each sample member was assigned to the first group for which he or she qualified. For example, if someone was both a dropout (Group 3) and was in a private school in 1988 (Group 5), he or she was assigned to Group 3.

The data used to assign the students to groups were drawn from a variety of possible sources, including questionnaire data for variables such as race and school sector type. If status at time of data collection was relevant and was not determined at the time of data collection, the imputed status developed during the NELS:88 second follow-up weighting process was used.

All sample members selected for inclusion in the sample were eligible to participate except for those who had died and those who were confirmed to be foreign exchange students at the time of the NELS:88 second follow-up interview and had returned to their country of origin by the time of the NELS:88 third follow-up survey.

3.7 Sample Design and Implementation: NELS:88 Fourth Follow-up Study

The sampling frame for the fourth follow-up study was constructed using all 15,964 persons or assumed persons or frame elements available from the sample frame for the third follow-up study—both respondents and nonrespondents to the third follow-up. Of these cases, 80 subjects were found to be ineligible (i.e., sample members who were deceased, permanently incapacitated, institutionalized, out of the country), and one subject was duplicated. These subjects were excluded, resulting in the fourth follow-up frame of 15,883 subjects.

Table 3.6-A.—Sample groups developed for the NELS:88 third follow-up study: 1994

Group	Group Description
0	*Excluded from NELS:88 third follow-up* The third follow-up sample was a spring-term-defined sample. Thus, students who were freshened into the sample and then dropped out by the time of data collection were added to this group. The group also included base-year dropouts and sample members who were ineligible or out of scope.
1	*Nonresponders* This group included sample members who had never completed a NELS:88 questionnaire in any round prior to 1994.
2	*Poor responders* These are sample members who did not complete either a second follow-up questionnaire or a questionnaire in their first eligible round.
3	*Ever dropped out* This group included sample members for whom investigators had evidence that they had ever dropped out of school (including those who were in school during periods of data collection).
4	*Ineligible to participate* This group included sample members who were ineligible for participation because of a language barrier or mental or physical disability prior to 1992.
5	*Attended a private school in 1988*
6	*Attended a private school in either 1990 or 1992*
7	*Hispanic*
8	*Asian/Pacific Islander*
9	*American Indian/Alaska Native*
10	*Black, top quartile in cognitive tests*
11	*Black, other test scores*
12	*White, lowest SES quartile*
13	*White, highest SES quartile*
14	*White, middle SES quartiles*
15	*Freshened in 1990*
16	*Freshened in 1992*
17	*Other*

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1992.

Table 3.6-B.—NELS:88 third follow-up sample groups and weighting sample results: 1994

Group	Selection Probability	Second Follow-up Sample				Third Follow-up Sample			
		n	Raw weight			n	Raw weight		
			Sum	Mean	Std Dev		Sum	Mean	Std Dev
Total		21,635	3,335,156	154	188	15,964	3,200,425*	200	226
0. Excluded	0	731	134,781	184	184	0	†	†	†
1. Nonresponders	0.15	288	56,688	197	258	43	56,720	1319	180
2. Poor responders	0.25	2,383	400,131	168	208	596	400,131	671	244
3. Dropouts	1	2,351	428,095	182	269	2,351	428,094	182	269
4. Ineligible	0.9	212	45,372	214	137	191	45,382	238	127
5. Private school, 1988	0.8	2,984	322,989	108	197	2,387	322,990	135	212
6. Private school, 1990–92	0.8	122	45,976	376	382	98	45,976	469	378
7. Hispanic	0.9	1,629	192,756	118	134	1,466	192,756	131	136
8. Asian/Pacific Islander	1	874	66,638	76	78	874	66,638	76	78
9. American Indian	1	132	21,457	163	105	132	21,457	163	105
10. Black, high test	1	79	13,545	171	134	79	13,545	171	134
11. Black, other	0.9	1,238	241,203	194	257	1,114	241,211	217	265
12. White, low SES	1	1,295	203,391	157	118	1,295	203,391	157	118
13. White, high SES	0.6	2,536	410,279	162	156	1,522	410,279	270	176
14. White, mid SES	0.8	4,763	749,524	157	134	3,810	749,524	197	138
15. F1 freshened	0.3	4	370	93	6	1	370	370	—
16. F2 freshened	0.3	6	690	115	59	2	690	345	—
17. Other	0.4	8	1,271	159	84	3	1,271	424	—

† Not applicable.

*Target total weight for 1994 was the total of 1992 sample weights less Group 0, or 3,200,375.

— Too few cases for reliable estimate.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1994.

The NELS:88 fourth follow-up study was the study team's first contact with the cohort in almost six years (eight years for sample members who did not respond in the third follow-up in 1994). The sample cohort members were approximately 26 years old, a time during their lives when young adults are very mobile. Indeed, in the six years prior to the fourth follow-up, many sample members had changed their addresses, and many had migrated to other areas of the country for purposes of education, employment, and family. Thus, as noted later, tracing and locating sample members was a very important part of the study. Efforts to reduce tracing and other data collection costs, while maintaining high sample response rates, encouraged the study team to implement a subsample before the start of data collection for the fourth follow-up to reduce the numbers of sample members with the most outdated locator information.

Ultimately, subsampling in the third follow-up was motivated by the desire to control data collection costs and schedules while, at the same time, ensuring adequate representation of specific domains or subpopulations in the subsample and preserving satisfactorily high response rates. To this end, subjects were classified into strata that reflected the subjects' response history through the third follow-up and identified specified domains of interest. Subsampling rates were then assigned to each of the strata. Strata defining small domains and domains having high analytic importance were assigned high subsampling rates, ranging from 0.60 to certainty. Lower subsampling rates (from 0.15 to 0.30) were assigned to those strata that identified subjects with poor response rates.[37]

As a consequence of the subsampling, substantial unequal weighting effects were introduced resulting in concomitant increases in the sampling variances of estimates of parameters that describe domains of subjects belonging to different strata. The motivation and objectives for subsampling in the fourth follow-up were the same as those for the third follow-up study, with the additional objective of controlling the effects of further unequal weighting.

With these objectives in mind, two competing optimization models were evaluated. The first model minimized a function describing the variable survey costs with constraints imposed on the overall unequal weighting effect, weighted overall response rate, and minimum sample sizes in specified domains. The second model minimized the overall unequal weighting effect subject to constraints imposed on the weighted overall response rate and minimum domain sizes. The numerical optimization routine manipulated the subsampling rates for the third follow-up strata given expected response rates and per unit costs to minimize the cost function or the overall unequal weighting effect, respectively.

By alternating between the two models, solutions were obtained that were recognizably better than those provided by either approach alone. Table 3.7-A summarizes the resulting subsample allocation. Based on these results, an initial subsample size of 15,236 subjects was selected, yielding 14,900 third follow-up study respondents and 336 nonrespondents of all types, including 16 hostile refusals. The subsample was selected with probability proportional to size, with the third-follow-up weights providing the measures of size. This procedure produced less variation in the weights than would be obtained, for example, by a stratified simple random sample.

Toward the end-stage of the data collection period, the researchers also implemented a second subsample to ensure adequate response rates among key respondent groups that would meet existing NCES data collection standards. Actual fourth follow-up response rates for dropouts and for the key

[37] As might be expected, sample members with low response rates are closely related to higher expected data collection costs than groups with higher response rates.

Table 3.7-A.—Pre-data collection subsample allocation for the NELS:88 fourth follow-up study: 2000

F3 Group	Locator Stratum													
	Respondent in F3						Hostile refusal		Other nonrespondent status in F3					
	4S*	4N	3S	3N	12S	12N	1S	1N	4S	4N	3S	3N	12S	12N
Nonrespondents or poor respondent to F3 and earlier data collections														
Eligible for subsample	48	7	169	94	60	75	4	16	0	0	2	1	12	116
Selected for subsample	48	7	169	94	60	75	0	0	0	0	1	0	4	34
Design sampling rate	1.00	1.00	1.00	1.00	1.00	1.00	0.05	0.05	1.00	1.00	0.35	0.20	0.35	0.30
Actual sampling rate	1.00	1.00	1.00	1.00	1.00	1.00	0	0			0.50	0	0.33	0.29
Other sample group for the third follow-up study														
Eligible for subsample	3919	52	7899	207	2296	74	103	35	18	2	188	91	253	142
Selected for subsample	3919	52	7899	207	2296	74	14	2	18	2	77	26	121	37
Design sampling rate	1.00	1.00	1.00	1.00	1.00	1.00	0.15	0.05	1.00	1.00	0.60	0.30	0.60	0.30
Actual sampling rate	1.00	1.00	1.00	1.00	1.00	1.00	.14	.06	1.00	1.00	.41	.29	.48	.26

* The numeric characters refer to the sample member's status on the fourth follow-up study's mailing to sample members and their contacts. The leading character, 4, indicates an address update card was returned with a new or confirmed address; 3 indicates the card was not returned; 1 and 2 indicate the card was returned as undeliverable or that no address was available for the initial mailing. No card was sent to F3 sample members classified as hostile refusals. The last character refers to the availability of a sample member or parent Social Security Number (S = available, N = not available).

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

racial-ethnic domains (i.e., Asian/Pacific Islander, American Indian/Alaska Native, Hispanic, non-Hispanic Black, non-Hispanic White) were examined, and these domains were used as strata to control the distribution of a further subsample. The objectives of the further subsampling were to achieve an overall weighted response rate of at least 84 percent and weighted response rates of at least 70 percent for the selected domains of interest.

The fourth follow-up nonrespondents were classified into three groups based on the likelihood of obtaining an interview. Within each group, sample members' weights were ranked and the resulting ordered array partitioned into three weight groups to differentiate among very large, very small, and "other" or medium weights. A formal nonlinear optimization[38] was carried out to minimize the overall unequal weighting effect subject to the overall and domain-level response rate objectives. The subsample size solutions, summarized in table 3.7-B, directed the majority of the subsample into the higher response categories, as desired.

Table 3.7-B.—Marginal subsampling rate solutions for the NELS:88 fourth follow-up nonresponse subsample: 2000

Stratum Margin	Sampling Rate
Overall	0.130
Weight Group Margin	
Small weight	0.150
Medium weight	0.116
Large weight	0.154
Response Group Margin	
Low response	0.014
Medium response	0.010
High response	0.557

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Overall, the frame constructed for the further subsampling activity accounted for 1,660 individuals from which 379 subjects were selected. Of these, 219 respondents were obtained and 9 ineligible subjects were identified. As before, the sample was selected with probability proportional to size, with the nonrespondents' weights providing the size measure.

3.8 Sample Weights

This section provides a description of the sample weights estimated in the five rounds of NELS:88 and outlines the procedures used to calculate those weights. Key statistical properties are also provided for each weight. Section 3.8.1 describes the base-year study in 1988; Section 3.8.2 covers the two high school follow-up rounds in 1990 and 1992; Section 3.8.3 deals with the third follow-up study in 1994; and Section 3.8.4 describes the fourth follow-up study. Researchers will find these weights on the

[38] Convergence criteria for the optimization routine were based on achieving the Karush-Kuhn-Tucker necessary conditions to an arbitrary degree of numerical accuracy. For additional information, see Chong and Zak (1996).

three public-use and four restricted-use ECBs: 1988-92 weights on the base-year to second follow-up ECBs; 1994 weights on the third follow-up ECB; and 2000 weights on the fourth follow-up ECB. A special restricted-use ECB containing postsecondary transcripts and associated weights will be available on a fourth ECB.

The general purpose of weighting survey data is to compensate for unequal probabilities of selection and to adjust for the effects of nonresponse. Weights are often calculated in two main steps. In the first step, unadjusted weights are calculated as the inverse of the probabilities of selection, taking into account all stages of the sample selection process. In the second step, these initial weights are adjusted to compensate for nonresponse; such nonresponse adjustments are typically carried out separately within multiple weighting cells. This weighting process was applied to all rounds of the NELS:88 data.

3.8.1 Base-Year Sample Weighting Procedures

The base-year weights were based on the inverse of the probabilities of selection into the sample and on nonresponse adjustment factors computed within weighting cells. Two different weights were calculated to adjust for the fact that not all sample members have data for all instruments. The base-year questionnaire weight (*BYQWT*) applies to 24,599 student questionnaires (and is also used in conjunction with parent data), while *BYADMWT* applies to the 1,035 school administrator questionnaires (17 base-year school principals failed to complete a school questionnaire). These weights project to the population of approximately 3,008,080 eligible 8th-graders in public, Catholic, and other private schools during the 1987-88 school year. The base-year weighting procedures consisted of two basic stages:

1. Calculating a preliminary base-year weight based on the inverse of the product of the probabilities of selection for the base-year sample.

2. Adjusting this preliminary weight to compensate for "unit" nonresponse, that is, for noncompletion of an entire school questionnaire or student questionnaire. The unit varied depending on the weight being adjusted.

School Design Weights. The first step in estimating preliminary base-year weights was to calculate a school design weight, *SCHWT*, by taking the reciprocal of each school's selection probability:

$$SCHWT_i = 1 \Big/ P_{il},$$

where P_{i1} is the selection probability for the i^{th} school.

To calculate P_{i1}, the research team first estimated the unconditional probability that a school was selected into Pool 1 or 2. Unconditional probability means that a school's chance of selection was not conditioned on the event that it was or was not selected into the NAEP sample. For schools selected into the core sample, the conditional probability of selection into NELS:88 given selection into NAEP was zero. Thus, for NELS:88 core schools, the unconditional probability of selection in NELS:88 was the product of the following two factors: P_{ci1}, the conditional probability of selection in NELS:88 given nonselection into NAEP, and $1\text{-}P_{ni1}$, where P_{ni1} is the probability of selection into NAEP. P_{ui1}, the unadjusted probability of selection in NELS:88, was obtained as follows:

$$P_{uil} = P_{cil} \times (1 - P_{nil}).$$

P_{ni1}, the probability of selection into NAEP, was not known for most of the schools and had to be estimated. Westat, Inc., the organization that selected the NAEP sample, provided the NELS:88 research team with the NAEP selection probabilities for the schools that were selected into NAEP. However,

Westat did not know and could not provide NAEP selection probabilities for the schools that were selected into NELS:88. To estimate the latter probabilities, regression analyses were used to predict the NAEP selection probabilities from school variables that were in the sampling frame. The predictor variables used were the number of students enrolled in the school, the estimated number of students in the 8th grade, the type of school (public, Catholic, or other private), and the percentages of students who were White, Black, and Hispanic.

With known values of P_{ci1} and with estimated values of P_{ni1}, the research team estimated the unconditional selection probabilities for all schools that were eligible for the NELS:88 sample. Two sets of probabilities were computed, one for the core sample plus private schools in one state augmentation sample, and another for the core sample plus all of the state augmentation samples. The former is the set used for the weights in the national public-use file. The latter set of probabilities is used for weights for all state augmentation samples and for estimating response propensities for schools (discussed below). The results of the regression were tested against subsamples of schools for which NAEP probabilities were known.

To smooth out the possible effects of errors in the estimates of the NAEP selection probabilities, the research team multiplied the unconditional selection probabilities by factors in each stratum to force their sum in the stratum to equal the number of schools that were sampled (i.e., that NORC attempted to contact) from that stratum. Thus, $P_{ui1\text{-}adj}$, the adjusted unconditional selection probabilities, were calculated as:

$$P_{ui1-adj} = P_{ui1} n_i \Big/ \sum_{jS(i)} P_{uj1},$$

with $\sum_{jS(i)}$ denoting summation over all schools j in the stratum to which i belonged, and n_i denoting the number of schools sampled from the stratum.

Statisticians then calculated P_{i1} according to the formula:

$$P_{i1} = P_{ui1-adj} \times F_{i1},$$

with $P_{ui1\text{-}adj}$ defined as the adjusted conditional probability that the school was selected into Pool 1 or 2, and F_{i1} defined as the fraction of schools in Pools 1 and 2 for the school's stratum that the team attempted to include in the survey. Taking the reciprocals of the selection probabilities yielded the sample design weights for the schools:

$$SCHWT_i = 1 \Big/ \left(P_{ui1-adj} \times F_{i1}\right).$$

Nonresponse-Adjusted Weights for Schools. Nonresponse-adjusted weights were derived as the product of the school's sample design weight multiplied by a nonresponse adjustment factor. Initial approximations to the nonresponse adjustment factors were calculated by first using linear and nonlinear logistic regression to estimate a propensity function, which gives the school's conditional probability of participation expressed as a function of school characteristics. The regression-based propensity function approach was used rather than the traditional weighting cell approach in order to include a number of variables in the adjustment process while avoiding the problem of small cells. Each school's design weight was divided by its estimated propensity. These first approximations were multiplied by factors so that the products would sum to known totals for the superstrata.

When estimating the propensity function, it is important to have available a set of variables that correlate well with participation in the survey. In many surveys, data necessary to accurately estimate propensities are either severely limited or unavailable. For NELS:88, the study team conducted a special survey of nonparticipating schools in Pool 1, in which a small selection of key items from the school questionnaire were asked in order to obtain data to estimate propensities. This sample will be referred to as the "nonparticipant" sample in the following description. The nonparticipant sample and the sample of responding schools were combined, and a dummy variable representing participation was created such that the nonparticipant schools were coded as "1" and the responding schools were coded as "0". This variable was used as the dependent variable in regression analyses used to estimate the propensity to nonrespond. The nonparticipant survey provided a basic set of descriptive information about nonresponding schools that, combined with the same information on responding schools, could be used as a set of independent variables in the regression analyses for estimating propensity to nonrespond.

To estimate the propensity function, stepwise linear regression was used to choose a subset of variables that correlated well with participation. Next, logistic regression was used to fit the propensity function. Once the logistic regression function was estimated, propensity estimates were produced for all of the schools for which school questionnaires and student questionnaires were available. For a small percentage of schools (about 2 percent), the research team obtained student data but was unable to obtain school data. For these schools, propensity estimates were calculated for the construction of the nonresponse-adjusted school weight, *BYADMWT,* which was used for the construction of weights for students and parents. The propensity estimates for these schools were derived from a reduced regression model that used only variables that were available from the sampling frame. The reduced model included school type (public, Catholic, and other private), urbanicity (urban/suburban/rural), geographic division (based on the nine Census regions as of 1988), and the estimated number of students in the 8th grade class. In addition to those variables, the full model included an indicator of whether entrance tests were used as a criterion for acceptance into school, and a measure of the frequency with which standardized test results were provided to the family. The propensity estimates were constrained to be at least 0.4, so that their reciprocals did not exceed 2.5.

Dividing *SCHWT* by the appropriate estimated propensity yielded a preliminary approximation to *BYADMWT:*

$$\mathrm{BYADMWT}_{\mathrm{prelim,1}} = \mathrm{SCHWT}_{\mathrm{i}} \Big/ \mathrm{PROP}_{\mathrm{i1}},$$

where $PROP_{i1}$ is the estimated propensity for school *i*.

The final weight was developed by multiplying the preliminary approximation by a factor that was constant within, but varied across, superstrata. The factor was chosen to ensure that for each superstratum the sum of $BYADMWT_{prelim}$ multiplied by an estimate of the 8th grade enrollment, *Y*, over all schools with school questionnaires was equal to the sum of *Y* in that superstratum in the frame. Thus:

$$BYADMWT_i = BYADMWT_{pre\lim_i} * \sum j \in S(i) Y_j \,/ (\sum j \in S(i) Yj * BYADMWT_{pre\lim,j} * PAR_{j1})$$

with

PAR_{j1} = 1, if school *j* participated, and

PAR_{j1} = 0, otherwise,

and $\sum_{jS(i)}$ denotes summing over all schools *j* in the stratum *i* to which school *i* belongs.

Second-Stage Sample Design Weight for Students. The second-stage sample design weight for students, *RAWWT,* is equal to the reciprocal of P_{i2}, the conditional probability that the student was selected given that his or her school was selected into the base-year sample, that is:

$$\text{RAWWT}_i = 1 / \text{P}_{i2} .$$

Student Selection Probabilities. Within each selected 8th grade school, rosters of all 8th grade students were obtained by the interviewer. At the time this list was prepared, the interviewer was also asked to classify each student into three groups: (1) Asian/Pacific Islander, (2) Hispanic, (3) all others. The rosters were used as within-school sampling frames, and ethnic classification was used in the oversampling of students of Asian/Pacific Islander and Hispanic descent.

Researchers used the initial roster and classification to construct three separate lists of students who were designated either as Asian/Pacific Islander, Hispanic, or non-Asian/Pacific Islander and non-Hispanic. These three lists were separately sampled with selection probabilities determined as follows:

1. Subject to the two upper bounds discussed below, students designated as Asian/Pacific Islander in the i^{th} school were sampled at a rate equal to $0.054/\text{p}_{i1}$, where p_{i1} is equal to the probability of selection for the i^{th} school (the same as F in the equations in Section 3.2.4), and 0.054 is the empirically-derived proportionality constant.

   This probability, $0.054/\text{p}_{i1}$, was subjected to the following upper bounds prior to its application. First, it was bounded at unity (1.0), and, second, it was bounded by a cap at 25 on the number of Asian/Pacific Islander students that would be selected in a sample school. Thus, the sampling rate for Asian/Pacific Islander students was set at
   $p_{ia2} = \min \{0.054/\text{p}_{i1}, 1, 25 / \text{N}_{ai}\}$, with N_{ai} defined as the number of eligible Asian/Pacific Islander students in school *i.*

2. Subject to the two upper bounds, students designated as Hispanic in the i^{th} school were sampled at a rate equal to $0.035/\text{p}_{i1}$, where p_{i1} is equal to the probability of selection for the i^{th} school, and 0.035 is the empirically derived proportionality constant.

   This probability, $0.035/\text{p}_{i1}$, was selected to the following upper bounds prior to its application. First, it was bounded at unity (1.0), and, second, it was bounded by a cap at 25 on the number of Hispanic students who would be selected in the sample school. Thus, the sampling rate for Hispanic students was set at $p_{ih2} = \min \{0.035/\text{p}_{i1}, 1, 25 / \text{N}_{hi}\}$, with N_{hi} defined as the number of eligible Hispanic students in school *i.*

3. Students designated as non-Asian/Pacific Islander and non-Hispanic (i.e. other) in the i^{th} school were sampled at a rate equal to

   $$p_{io2} = 24/\text{TSIZE}_{i1},$$

   where $TSIZE_i$ is equal to the total number of 8th-graders not pre-identified as Asian/Pacific Islander or Hispanic on the roster for the i^{th} school.

One additional step in the process of producing student weights involved weighting that is linked to the selection of students within sample schools. In this step, the weight factor for each student was equal to the inverse of the student's probability of selection into the sample within the sample school. For example, if within a certain school, a selected student received a probability of selection equal to 1/20 (or 0.05), the student's corresponding weight would be equal to 1/.05 = 20.0.

It should be noted that a student's probability of selection was determined by the initial classification (Asian/Pacific Islander, Hispanic, or other) that the student was given at the time of selection. In those situations where the initial selection was incorrect, the probability of selection for the student was equal to the selection probability actually used, rather than to a theoretical probability under the assumption of perfect classification.

Asian/Pacific Islander and Hispanic Oversamples. As part of the overall design of NELS:88, Asian/Pacific Islander and Hispanic students were oversampled in order to ensure adequate sample sizes for subgroup analyses. This oversampling was implemented by increasing the probability of selection at the within-school stage of the selection process.

Nonresponse-Adjusted Student Weights. The basic nonresponse-adjusted student weight, *BYQWT*, was derived as the product of the student's sample design weight (*RAWWT*) multiplied by a nonresponse adjustment factor. The factor was intended to adjust for the fact that some of the sampled students did not participate, that is, did not return a completed questionnaire. To derive the nonresponse adjustment factor, NORC used a weighting-cell approach. First, the group of all sampled students was partitioned into weighting cells such that each sampled student belonged to exactly one cell. Statisticians attempted to construct cells so that students in the same cell had similar propensities to participate in the survey. Next, the nonresponse adjustment factor for a student in a given cell was computed as the ratio of the sum of *BYADMWT* (the nonresponse-adjusted weight for schools) multiplied by *RAWWT* for all students in the cell, to the sum of *BYADMWT* $\times$ *RAWWT* for all of the students who participated. The reciprocal of this factor may be interpreted as an estimate of the participation propensity for students in the cell because it is simply the weighted proportion of students who did participate. Thus, the nonresponse adjustment factor, *FAC,* for weighting class *c* is defined by

$$\mathrm{FAC_c} = \sum_{iec} \mathrm{BYADMWT_i} \times \mathrm{RAWWT_i} \,/ \sum_{\mathrm{iec}} \mathrm{BYADMWT_i} \times \mathrm{RAWWT_i} \times \mathrm{PAR_{i2}},$$

where $\sum_{iec}$ denotes summation over all students in the sample in weighting class *c,* and

PAR_{i2} = 1, if student *i* participated, and
PAR_{i2} = 0, otherwise.

$BYQWT_{prelim,i}$, the preliminary nonresponse-adjusted student weight for student *i,* is defined by

$$\mathrm{BYQWT_{prelim,i}} = \mathrm{BYADMWT_i} \times \mathrm{RAWWT_i} \times \mathrm{FAC_{ci}},$$

where *ci* denotes the weighting class to which the student belongs.

The cells were based on classification of the students according to data that were available from the school rosters and from the sampling frame for the schools. The cells were set up as shown in table 3.8.1.

Classification by school type and region was based on information included in the sampling frame. Ethnicity was obtained when rosters were collected from schools. Gender was indicated on some but not all of the rosters; where it was not indicated, it was inferred on the basis of the students' first names. Comparison of the inferences with responses to the questionnaires showed a high degree of

Table 3.8.1.—Weighting cells used for nonresponse adjustment of NELS:88 base-year student weights: 1988

School Type	Region	Ethnicity	Gender
Public	Northeast	Other*	Male Female
	Mid-Atlantic	Other	Male Female
	Other	Other	Male Female
		Asian/Pacific Islander	Male Female
		Hispanic	Male Female
Private	Any	Other	Male Female
		Asian/Pacific Islander	Male Female
		Hispanic	Male Female

*The ethnicity of "Other" means not Hispanic and not Asian/Pacific Islander.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

accuracy. In the weighting cell information, however, the questionnaire data for gender were not used even when available, so that the classification for participants and nonparticipants was consistent.

To obtain the final nonresponse-adjusted student weight, the nonresponse adjustment factor was modified by a "polishing" factor. The polishing factors were defined for eight classes of students, four types of schools by two sexes. The polishing factor for a class was set equal to the ratio of the sum of $BYADMWT \times RAWWT$ for all students in the class who participated. Polishing preserves the sums of the weights across the eight classes. The polishing factor for any one of the eight classes of students, as class k, is POL_k, and is defined by:

$$\mathrm{POL_k} = \sum_{iek} \mathrm{BYADMWT_i} \times \mathrm{RAWWT_i} \Big/ \left(\sum_{iek} \mathrm{BYQWT_{prelim,I}} \times \mathrm{PAR_{i1}} \right),$$

where Σ_{iek} denotes summation over all students in class k. Then, $BYQWT$ for student i is calculated as:

$$\mathrm{BYQWT_i} = \mathrm{POL_{ki}} \times \mathrm{BYQWT_{prelim,i}},$$

where ki denotes the polishing class to which student i belongs.

3.8.2 Estimation of Cross-Sectional and Panel Sample Weights for 1990 and 1992

In the waves that followed the base year of NELS:88, weights were estimated that allowed analysis of key panel and cross-sectional populations. The sections below describe these weights and the procedures used in their calculation.

3.8.2.1 First Follow-up Sample Weights

Two weights were developed for the overall NELS:88 first follow-up sample. The first, or basic weight, applies to all members of the first follow-up sample who completed a first follow-up questionnaire, regardless of their participation status in the base year. The basic weight (*F1QWT*) allows projections to the population consisting of all persons who were either in the 8th grade during the 1987-88 school year or in the 10th grade during the 1989-90 school year. Thus, this population encompasses both populations of prime analytic interest—the population of 1990 10th-graders (including those who were not 8th-graders in 1988) and the 1988 8th-grade population (excluding any additional 1990 10th-graders). By selecting the appropriate sample members, analysts can use this basic weight to make unbiased projections to the first of these populations (i.e., 1990 10th-graders). The second, or panel weight, applies to all members of the first follow-up sample with complete data from both rounds of the study. The panel weight (*F1PNLWT*) can be used to make projections to the other key analytic population—1988 8th-graders (excluding those ineligible for base-year data collection).

Basic First Follow-up Weight (F1QWT). Calculation of the basic weight required somewhat different procedures for the three groups of the full first follow-up sample: 1988 8th-graders deemed eligible for the base-year survey, 1990 10th-graders who were not in the 8th grade in 1988, and 1988 8th-graders who were deemed ineligible for participation in the base year but were considered eligible to participate in the first follow-up.

Eligible 1988 8th graders. With a few exceptions, those individuals who were eligible for the base-year survey and selected into the base-year sample in 1988 remained eligible for the first follow-up sample. The exceptions involved cohort members who died, left the country, or suffered grave impairments between 1988 and 1990.

The first step in constructing a basic weight for these sample cases involved developing a design weight (also called a raw weight) that reflected the selection probabilities for each case. Each student selected for the base-year sample (including base-year nonparticipants) was assigned a base-year design weight (*BYDW*) based on his or her probability of selection into the base-year sample. The base-year design weight reflected both the probability of selecting the base-year school (inflated to adjust for school-level nonresponse) and the probability of selecting the student given that the school had been selected and had agreed to participate. The base-year design weight does not adjust for student-level nonresponse. The base-year design weight was then multiplied by the inverse of the case's probability of selection for the first follow-up sample; the latter probability took into account the subsampling done during the first follow-up. More formally, the first follow-up design weight (*FFUDW*) for student *i* was defined as:

$$\mathrm{FFUDW_i} = \mathrm{BYDW_i} \times \left({}^{1}\!\!\diagup\!\!{}_{\mathrm{P_{1i}}} \right),$$

in which P_{1i} represents the probability of selection for the first follow-up sample.

The next step was to adjust the design weight for first follow-up nonresponse. Weighted response rates were computed for subgroups of this portion of the first follow-up sample. The weight used was the first follow-up design weight. The subgroups were

Out-of-sequence students (i.e., those who were not in 10th grade in 1990);

- Dropouts identified at the time of initial first follow-up sampling;[39]
- Students who had transferred out of the first follow-up school from which they were selected;
- Potential dropouts;
- Other students initially classified as attending schools with 3 or fewer base-year students; and
- Other students initially classified as attending schools with 4 or more base-year students.

The product of the inverse of the relevant response rate and the first follow-up design weight served as a preliminary adjusted weight. These preliminary weights were then further adjusted to meet overall and marginal targets for the sums of the weights. The target for a given marginal category was the sum of the final base-year weights for all base-year sample cases in that category. The categories were based on

- Base-year school type (public, Catholic, and other private[40]);
- Student sex (male and female);
- Race-ethnicity (non-Hispanic White, American Indian; Hispanic, Asian, non-Hispanic Black, and unknown); and
- Base-year school region (Northeast, Midwest, South, and West).

The preliminary adjusted first follow-up weights were further adjusted until the sum of the weights for each marginal category (e.g., males) was equal to the corresponding sum of the final base-year weights for that group. This final adjustment procedure is referred to as polishing, or multidimensional raking.[41] (See the Glossary, appendix E, for definitions of technical terms.)

1990 10th graders who were not 1988 8th graders. All members of this population who are included in the first follow-up sample were selected through the freshening process (see Section 3.4.1). This process linked each 1990 10th-grader who was not a 1988 8th-grader to a student who was an 8th-grader in 1988. The first follow-up design weight (*FFUDW*) for each student in the freshening sample is therefore equal to the first follow-up design weight of the base-year student to whom he or she was linked. For purposes of variance estimation, both students are considered members of the same stratum and school.

The nonresponse adjustment for this portion of the sample involved two steps. First, the first follow-up design weight (*FFUDW*) for responding students in the freshening sample was inflated by a factor equal to the inverse of the weighted response rate for this portion of the sample. (The first follow-up design weight was the weight used in computing this response rate.) Second, the marginal distributions of the weights of the respondents were adjusted, by raking, to match the corresponding

[39] See Section 3.4.1 for a discussion of the process of subsampling dropouts and potential dropouts for the NELS:88 first follow-up.

[40] The National Association of Independent Schools (NAIS) endorsed NELS:88. NAIS schools form a base-year school sampling stratum in NELS:88, and NAIS constitutes a category within the restricted-use file school control variable.

[41] Multidimensional raking was also used in the base-year weighting process. Although it is generally true that the base-year weight for a student should be less than the first follow-up weight, this relationship may sometimes be reversed. This is a consequence of the raking procedure. The use of raking may also sometimes produce a reversal of the ordering for panel weights (described in the next section) relative to the basic first follow-up weight; that is, the first follow-up panel weight for an individual may be less than the individual's basic first follow-up weight.

distributions for all cases selected through freshening (including nonrespondents). The two dimensions used in the raking procedure were sex and race-ethnicity (non-Hispanic White, American Indian, Hispanic, Asian, non-Hispanic Black, and unknown as the categories).

1988 Ineligible 8th graders who were eligible for the first follow-up. A number of students who were not capable of participating in the base year were eligible for participation in the first follow-up. *F1QWTs* for these students were calculated during the course of the second follow-up weighting process and were developed using second follow-up procedures.

The first follow-up design weight was obtained by dividing the base-year design weight by 0.42 to allow for the subsampling that was done for this group. Nonresponse adjustment cells were defined based on a combination of their base-year and first follow-up status (see Step 2 in previous section), gender, and race (Asian/Pacific Islander, Hispanic, other). Each respondent's first follow-up design weight was then multiplied by the inverse of the weighted response rate (using the first follow-up design weight) for his or her cell. This adjusted weight serves as the respondent's *F1QWT*.

First Follow-up Panel Weight (F1PNLWT). The panel weight was developed only for those cases that were selected for both the base-year and first follow-up samples and that provided complete data in both rounds. The same procedures used in developing the basic first follow-up weight for 1988 8th-graders selected for the base-year sample were applied to the subset of them for whom complete data were obtained in both rounds. As with the basic first follow-up weight, the target sum of weights for the panel weight was the sum of the final base-year weights for all base-year sample subjects who remained eligible for the first follow-up sample. The same six nonresponse adjustment subgroups and multidimensional raking procedures used in calculating the basic first follow-up weight were also used in calculating the panel weight.

3.8.2.2 Second Follow-up Sample Weights

Explanation of Weights. Nine weights were developed for inclusion on the second follow-up data files. They include

F2QWT This cross-sectional weight applies to all members of the second follow-up sample who completed a second follow-up questionnaire, regardless of their participation status in previous rounds. It allows projections to the population consisting of all persons who were either in the 8th grade during the 1987-88 school year, in the 10th grade during the 1989-90 school year, or in the 12th grade in the 1991-92 school year. By selecting the appropriate sample members with the flag *G12COHRT,* analysts can use *F2QWT* to make unbiased projections to such populations as 1992 12th graders.

F2PNLWT This panel weight applies to sample members who completed a questionnaire in 1988, 1990, and 1992 (all three rounds of NELS:88at that time). This can be used to make projections to the population of 1988 8th-graders.

F2F1PNWT This panel weight applies to all sample members who completed both a first follow-up and a second follow-up questionnaire, regardless of base-year status. This allows projections to the population consisting of persons who were in the 10th grade in 1990 or the 12th grade in 1992. By selecting appropriate sample members with the flag *F2F1PNFL*, analysts can use *F2F1PNWT* to make projections to such populations as 1990 10th-graders.

F2CXTWT This cross-sectional weight applies to students who attended the schools selected for inclusion in the second follow-up teacher and school administrator components and who completed a second follow-up questionnaire. The population was restricted to early graduates and students who were in the schools during spring data collection. This weight allows analysts to generate national student-level statistics using the teacher and school administrator data despite the bias against small cluster sizes in sample selection.

F2TRSCWT This cross-sectional weight applies to all early graduates, dropouts, students in sampled schools during spring data collection, and all sample members who were both ineligible for all three rounds of NELS:88 and were in the 12th grade during the 1991-92 school year for whom NORC received a transcript.

F2TRP1WT This panel weight applies to sample members who were participants in 1988, 1990, and 1992 (all three rounds of NELS:88) and for whom transcript data are available. *F2TRP1WT* allows analysts to perform panel analyses using transcript data in conjunction with 1988, 1990, and 1992 test and questionnaire data.

F2TRP2WT This panel weight applies to sample members who were participants in 1990 and 1992 (the first and second follow-ups) and for whom transcript data are available. *F2TRP2WT* allows analysts to perform panel analyses using transcript data in conjunction with 1990-92 test and questionnaire data.

F2PAQWT This cross-sectional weight applies to all students for whom NORC collected a parent questionnaire during the second follow-up.

F2EXPWT This cross-sectional weight allows licensed researchers to estimate dropout rates for grades 8 to 10, 10 to 12, and 8 to 12. In addition, the expanded sample weight assists researchers in the exploration of the impact of eligibility rules on survey estimates. Because of confidentiality requirements, the expanded sample file is available for use only by licensed researchers.

Process for Calculating Second Follow-up Weights. A basic four-step process was defined for the calculation of all nine questionnaire weights. The first step, developing a classification scheme, was performed at the beginning of the weighting process for all sample members. The values remained static and were used throughout the process for all weights. Steps 2 through 4 were followed for all weights, but the results of each were tailored according to the characteristics of each weight's specific population.

Step 1. Develop a classification scheme. All sample members were divided into basic sample groups depending upon their status during data collection for each of the three rounds of NELS:88. Freshened students were assigned the status of their linked student for those rounds where they had not been in the sample. Students for whom status was unknown had their status imputed based upon the distribution of status across others in their base-year, first follow-up, or second follow-up categories, and where group size permitted, race and gender were also considered.

The eight basic classification categories for a single round are defined as

1. Eligible, dropout as of survey date;
2. Eligible, in school, in expected grade;

3. Eligible, in school, not in expected grade;
4. Ineligible
    a. In school, in expected grade,
    b. In school, not in expected grade,
    c. Not in school;
5. Out of scope (deceased or out of country);
6. Eligible, freshened, dropout as of survey date;
7. Eligible, freshened, in school; and
8. Ineligible, freshened.

In this classification scheme, "Dropout" (following the HS&B definition) generally refers to a student who has left a diploma-granting high school program. This included members who were not pursuing an education at all, home-study students, members who were continuing their education in a nontraditional setting (e.g., preparing for the GED examination), and institutionalized sample members. There are two exceptions to this general rule. First, early graduates were included in the "in school" category. Second, because sample members who attended nontraditional schools during the first follow-up were classified as students then, they were treated as such during the calculation of their first follow-up status.

"Ineligible" refers to members who were not given the questionnaires because of a language barrier or a mental or physical incapacity.

"Expected grade" means 10th grade in the first follow-up and 12th grade or early graduate in the second follow-up.

Step 2. Establish second follow-up design weight. The design weight reflects the selection probabilities for each case for a given population. Sample members may have multiple design weights that vary depending upon the weight that is being calculated. For the weights unaffected by school sampling (*F2QWT, F2PNLWT, F2F1PNWT*) and for the dropouts, early graduates, and ineligible 12th graders in *F2TRSCWT*, the design weight used is equal to the first follow-up design weight.[42] Second follow-up freshened students take on the first follow-up design weight of the student they were linked to in the freshening process. When sample members are included because of their association with a sampled school in *F2TRSCWT* and for all members in the *F2CXTWT* population, the design weight is equal to the first follow-up design weight divided by their school's second follow-up selection probability. For students represented in the parent sample, the calculation of *F2PAQWT* uses the first follow-up design weight divided by the parent's second follow-up selection probability.

Step 3. Adjust for second follow-up nonresponse. Nonresponse adjustment cells were based on combinations of the classification values from Step 1, as well as race (Hispanic, Asian/Pacific Islander, other, unknown), and gender for the members of that weight's population. The second follow-up design weight for each responding sample member was inflated by a factor equal to the inverse of the weighted response rate for their cell. This yielded the nonresponse-adjusted weight. This step was performed independently for each weight calculated. For second follow-up freshened students, the nonresponse-adjusted weight served as their final weight.

[42]Included in the transcript data files are approximately 90 students who were ineligible in all three rounds of NELS:88 and were seniors in 1992.

Step 4. Perform multidimensional raking. Sample members who were not freshened in the second follow-up had their second follow-up nonresponse-adjusted weight further adjusted through a raking step. The total sum of the weights and percentage distributions that were used in raking were developed as follows:

Targets were developed that used the second follow-up expanded sample weight. The second follow-up expanded weight is a weight that was calculated for every sample member in order to estimate national dropout rates.[43] It was used in developing the weighted population targets to ensure consistency in dropout rates derived when using questionnaire weights. These targets were calculated separately for each of the eight questionnaire weights and reflected the characteristics of each weight's inference population. Two types of target numbers were developed. The sum of expanded weights for a given questionnaire weight's inference population was used as the target total population for that questionnaire weight. Weighted frequency distributions using the expanded weights associated with a questionnaire weight's inference population were calculated for dropout rates between base year and first follow-up, dropout rates between first follow-up and second follow-up, first follow-up status (from Step 1) and second follow-up status (from Step 1).

Additional percentage targets were developed for raking using first follow-up weights. Calculated independently for each of the eight weights according to the characteristics of each inference population, these targets used *F1QWT* for sample members who had been eligible for the first follow-up questionnaire, or the first follow-up design weight for those who were not. Weighted frequencies calculated using these weights were used as target distributions. These target categories included race (White, Black, Hispanic, Asian/Pacific Islander, American Indian, unknown), gender, base-year school region, base-year school type, and base-year school urbanicity.

3.8.2.3 Results of Weighting

The statistical properties of each NELS:88 sample weight are outlined in table 3.8.2.3-A (cross-sectional weights), table 3.8.2.3-B (panel weights), and table 3.8.2.3-C (contextual weights) below. Analysts should note that compared to the base-year questionnaire weight (*BYQWT*), the first follow-up questionnaire (*F1QWT*) and panel (*F1PNLWT*) weights and the second follow-up cross-sectional (*F2QWT*) and panel weights (*F2PNLWT, F2F1PNWT*) are larger on average and more variable. This most directly reflects the effect of subsampling students at different rates depending on the number of other NELS:88 students with whom they were clustered in their first follow-up schools.

[43] For sample members not freshened in the second follow-up, the process involved using a multidimensional raking procedure to adjust the second follow-up design weight where the marginal target categories were based on roster race (Asian/Pacific Islander, Hispanic, other, unknown) and gender, base-year school type, base-year school region, base-year school urbanicity, and the status values from the classification scheme described above in Step 1. Target margins for the expanded weight were calculated using the first follow-up expanded sample weight (a similar weight developed in the first follow-up for estimating the 1988-90 dropout rate) for students for whom one was calculated and first follow-up design weights for the first follow-up sample members who did not receive a first follow-up expanded weight (such as freshened students). Second follow-up freshened students had their second follow-up design weight as their expanded sample weight. This step was performed for the sample as a whole.

Table 3.8.2.3-A.—Statistical properties for the NELS:88 base-year through fourth follow-up sample weights for all sample members, 1988-2000

Property	BYQWT	F1QWT	F2QWT	F3QWT	F4QWT
Mean	122.29	165.88	167.75	214.67	259.27
Variance	4,359.16	46,249.54	43,671.80	55,899.72	162,623.40
Standard deviation	66.02	215.06	208.98	236.43	403.27
Coefficient of variation (×100)	53.99	129.65	124.58	110.14	155.54
Minimum	2.44	2.14	2.14	7.96	8.71
Maximum	836.91	6,996.81	6,670.09	6,135.13	20,898.71
Skewness	2.18	10.89	10.18	7.65	16.68
Kurtosis	16.32	205.24	180.09	108.61	629.81
Sum	3,007,779	3,217,069	3,224,099	3,201.74	3,148,607.79
Number of cases	24,599	19,394	19,220	14,915	12,144

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-2000.

Table 3.8.2.3-B.—Statistical properties for the NELS:88 first and second follow-up panel weights for all panel members: 1990-1992

Property	F1PNLWT	F2PNLWT	F2F1PNWT
Mean	172.62	180.17	174.66
Variance	52,603.86	50,610.95	46,174.76
Standard deviation	229.36	224.97	214.88
Coefficient of variation (×100)	132.86	124.86	123.03
Minimum	2.26	2.39	2.31
Maximum	7,479.71	7,388.13	6,780.07
Skewness	11.22	11.59	10.63
Kurtosis	214.14	233.6	196.94
Sum	3,007,813	2,970,835	3,164,096
Number of cases	17,424	16,489	18,116

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1990-1992.

Table 3.8.2-C.—Statistical properties of the expanded sample, parent, and contextual weights for the NELS:88 second follow-up study: 1988-1992

Property	Expanded Sample Weight F2EXPWT	Parent Weight F2PARQWT	Contextual Weight F2CXTWT
Mean	155.64	196.41	171.77
Variance	36,432.58	61,981.79	102,513.57
Standard deviation	190.87	248.96	320.18
Coefficient of variation (×100)	122.64	126.76	191.05
Minimum	2.28	2.39	1.98
Maximum	5,255.58	7,307.02	12,025.09
Skewness	9.37	9.26	19.14
Kurtosis	147.52	151.42	543.71
Sum	3,289,133	3,247,182	2,695,994

NOTE: *F2TRSCWT* = high school transcript collected in the second follow-up; *F2TRP1WT* = participant in 1988, 1990, and 1992 and transcript collected in second follow-up; *F2TRP2WT* = participant in 1990 and 1992 and transcript collected in second follow-up.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

3.8.3 Weighting in the NELS:88 Third Follow-up (NELS:88/94)

The purpose of weighting in the 1994 round was once more to compensate for unequal probabilities of selection and to adjust for the effects of nonresponse. In the first step, unadjusted weights (raw or design weights) were calculated as the inverse of the probabilities of selection, taking into account all stages of the sample selection process. In the second step, these initial weights were adjusted to compensate for unit nonresponse to generate final or nonresponse-adjusted weights; such nonresponse adjustments are typically carried out separately within multiple weighting cells.

In order to maintain consistency in weights across the various waves and across the various weights within waves, multidimensional raking was also applied when creating NELS:88 weights. In the third follow-up, raking was performed with respect to base-year school characteristics, race, gender, and status in each of the rounds.

3.8.3.1 Calculation of Third Follow-up Weights

The following procedures were used to calculate the weights for use with the third follow-up data. Eleven different weights, each designed to reflect a different population of respondents and to be used for different analyses were calculated (see descriptions below). (Note that while the following weights appear on the 1988-94 restricted-use ECB, several of the cross-sectional weights are absent from the public-use ECB.) These weights are not repeated on the fourth follow-up ECB and are described here for information purposes.

F3QWT This weight applies to all members of the third follow-up sample who completed a questionnaire in 1994, regardless of their participation status in previous rounds. When used with the appropriate sample flags (*F3UNIV2A, F3F1STFL*, and *F3UNIV2D*), this weight allows projections to the following populations:

Spring 1988 8th-graders eligible to complete questionnaires in 1992 and 1994, regardless of 1988 and 1990 eligibility;

Spring 1990 10th-graders eligible to complete questionnaires in 1992 and 1994, regardless of 1990 eligibility; and

Spring 1992 12th graders eligible to complete questionnaires in 1992 and 1994.

F3PNLWT This panel weight applies to sample members who completed questionnaires in all four rounds of NELS:88 up to and including the third follow-up study. *F3PNLWT* can be used in longitudinal analyses to make projections to the population of spring 1988 8th-graders.

F3F1PNWT This panel weight applies to sample members who completed questionnaires in 1990, 1992, and 1994, regardless of base-year status. *F3F1PNWT* allows projections (when used with the flag variable *F3F1PNFL*) in longitudinal analyses to the population of spring 1990 10th-graders.

F3F2PNWT This panel weight applies to sample members who completed questionnaires in 1992 and 1994, regardless of base-year or first follow-up status. *F3F2PNWT* allows projections (when used with the flag variable *F3F2PNFL*) in longitudinal analyses to the population of spring 1992 12th graders.

F3CXTWT This weight is to be used in analyzing the 1994 sample using the 1992 school administrator and teacher data. It applies to 1994 respondents who were early graduates from or students in the spring of 1992 at contextual sample second follow-up schools and who completed a 1992 questionnaire. (Teacher and school administrator data were collected from a contextual school subsample of the 1992 schools.) This weight allows analysts to generate national statistics for 1994 respondents using the associated 1992 teacher and school administrator data despite the bias against small cluster sizes in sample selection.

F3PAQWT This weight is to be used in analyzing the 1994 sample in conjunction with the 1992 parent data. It applies to all 1994 respondents for whom second follow-up parent questionnaire data were collected.

F3TRSCWT This weight is intended to be used in analyzing the 1994 sample with the high school transcript data collected in the second follow-up. It applies to 1994 respondents whose spring 1992 status was dropout, early graduate, or student in a sampled school and for whom transcripts were collected in 1992.

F3QWTG8 This weight is equal to *F3QWT* for 1994 respondents who were in the 8th grade in the spring of 1988 and is equal to zero for all other respondents. Use of this weight allows projections to the population of spring 1988 8th-graders who were eligible to complete questionnaires in 1992 and 1994, regardless of 1988 and 1990 eligibility.

F3QWTG10 This weight is equal to *F3QWT* for 1994 respondents who were in the 10th grade in the spring of 1990 and is equal to zero for all other respondents. For this weight, 1990 10th grade cohort membership is based on the 1990 enrollment status used in 1994 weighting. For sample members whose status was not determined in 1990, 1990 enrollment status was imputed. *F3QWTG10* allows projections to the

population of spring 1990 10th-graders who were eligible to complete questionnaires in 1992 and 1994, regardless of 1990 eligibility.

F3QWTG12 This weight is equal to *F3QWT* for 1994 respondents who were in the 12th grade in the spring of 1992 and is equal to zero for all other respondents. For this weight, 1992 12th grade cohort membership is based on the 1992 enrollment status used in 1994 weighting. For sample members whose status was not determined in 1992, 1992 enrollment status was imputed. *F3QWTG12* allows projections to the population of spring 1992 12th graders who were eligible to complete questionnaires in 1992 and 1994.

F3QWT92G This weight is equal to *F3QWT* for 1994 respondents who received a high school diploma between September 1, 1991, and August 31, 1992, or respondents whose diploma receipt date is not known but who began their postsecondary education between June 1 and October 31, 1992. *F3QWT92G* is zero for all other 1994 respondents. *F3QWT92G* allows projections to the population of persons who received a high school diploma in the 1991-92 academic year.

3.8.3.2 Process for Calculating Third Follow-up Weights

Expand the Second Follow-up Classification Scheme. As a part of the second follow-up weighting process, all sample members were divided into basic sample groups depending on their status during data collection for each of the three rounds of NELS:88. Freshened students were assigned the status of their linked student for those rounds where they were not yet in the sample. The possible values included

1. Eligible, dropout as of survey date;

2. Eligible, in school, in expected grade;

3. Eligible, in school, not in expected grade;

4. Ineligible,
   a. In school, in expected grade,
   b. In school, not in expected grade, or
   c. Not in school;

5. Out of scope (deceased or out of country);

6. Eligible, freshened, dropout as of survey date;

7. Eligible, freshened, in school; and

8. Ineligible, freshened.

Sample members for whom status was unknown had their status imputed based on the weighted distribution of status across others in their base-year, first follow-up, and second follow-up categories, and where group size permitted, race and gender were also considered.

In this classification scheme, "dropout" generally refers to a student who has left a diploma-granting high school program. This would include members who are not pursuing an education at all,

home-study students, members who are continuing their education in a nontraditional school, and institutionalized members. There were two exceptions to this general rule. First, early graduates were included in the "in school" category. Second, because sample members who attended nontraditional schools during the first follow-up were classified as students at that time, they were treated as such during the calculation of their first follow-up status. "Ineligible" refers to members who were not given the questionnaire due to a language barrier or a mental or physical disability. "Expected grade" means 10th grade in the first follow-up and 12th grade in the second follow-up.

A third follow-up status was defined and used in conjunction with the status categories developed during the second follow-up. The possible values for the third follow-up status included:

1. Eligible, received high school diploma;
2. Eligible, received GED or certificate;
3. Eligible, working toward high school diploma or equivalent;
4. Eligible, did not finish high school and is not working toward diploma or equivalency; and
5. Deceased or ineligible for third follow-up.

Sample members for whom status was not determined in 1994 had their status imputed using the method employed in the second follow-up. "Ineligible for third follow-up" refers only to sample members who were not given the questionnaire because they entered the NELS:88 sample as exchange students and had returned to their home country prior to the 1994 data collection.

Calculate the Third Follow-up Design Weight. The basic 1994 design weight was calculated at the time of the 1994 sampling. Sampling groups were defined, and each was assigned a percentage of cases to be selected. Cases were selected such that the overall selection probability was a fixed percentage per sampling group, but with the probability of selection within the group proportional to the second follow-up design weight. This design weight, *F3RAWWT,* was used to compute *F3QWT, F3F2PNWT, F3F1PNWT*, and *F3PNLWT*. *F3QWTG8, F3QWTG10*, *F3QWTG12,* and *F3QWT92G* were in turn derived from *F3QWT*. Using a similar procedure as was used during the second follow-up, the design weight used for *F3PAQWT* was *F3RAWWT* divided by the parent's second follow-up selection probability. The design weights for *F3TRSCWT* and *F3CXTWT* were *F3RAWWT* divided by the second follow-up school selection probability for those sample members whose inclusion was determined by school affiliation, or *F3RAWWT* for those who were included despite their school affiliation.

Calculate Third Follow-up Expanded Weight. This cross-sectional weight was developed for all members of the NELS:88/94 sample, regardless of their questionnaire completion status and was used to develop targets for the 1994 respondent weights. A multidimensional raking procedure was used to adjust the basic third follow-up design weight, *F3RAWWT*, where the marginal target categories were based on roster race (Asian/Pacific Islander, Hispanic, other) and gender, base-year school type, base-year school region, base-year school urbanicity, and the status values from the classification scheme described in above. Target margins were developed using the first follow-up expanded weight for students who were respondents to that wave; the second follow-up design weight was employed with students freshened into the sample.

For this weight only, the NELS sample members who were excluded from the third follow-up sample because they were deceased or ineligible for the second follow-up sample were included. This was to ensure a consistency in the population sizes across the rounds. These cases were dropped when

the targets were developed, thereby automatically shrinking the targets to accommodate the loss of the corresponding population members.

Adjustment for Nonresponse. Nonresponse adjustment cells for each 1994 weight were created based on combinations of the classification scheme described below, as well as roster gender and roster race (Hispanic, Asian/Pacific Islander, other) for the members of that weight's population. The steps for creating the nonresponse cells and adjusted weight included the following:

1. Cells were initially defined by dividing sample members into groups based on their base-year, first follow-up, and second follow-up status. Cells that had fewer than 50 members or fewer than 10 respondents were combined at the second follow-up level. Base-year and first follow-up distinctions were maintained, but within these, some relatively small cells were combined as necessary to achieve the minimum cell size.

2. Cells that contained more than 100 members and 20 respondents might have been eligible for division. A cell was divided if all resulting subgroups met the minimum 50/10 requirement. Divisions were first considered on the basis of third follow-up status, then roster gender, then roster race.

3. Once the cells were defined for a given weight, the appropriate third follow-up design weight for each responding member was inflated by a factor equal to the inverse of the weighted response rate for the cell.

Multidimensional Raking. Using *F2QWT,* targets were developed for each weight for race-ethnicity (White, Black, Hispanic, Asian/Pacific Islander, American Indian, other), gender, base-year school region, base-year school urbanicity, and base-year school type. Targets were developed for current and prior round status and total population sums for each weight using *F3EXPWT*.

The following table (table 3.8.3.2) includes summary statistics for each of the third follow-up weights.

3.8.4 Weighting in the NELS:88 Fourth Follow-up (NELS:88/2000)

In the absence of nonresponse and other nonsampling errors, unbiased estimates of parameters described by linear statistics are computed as weighted functions of response variable values. In particular, the weights are the inverses of the expected frequencies[44] with which population units appear in conceptually repeated samples selected using the sampling design developed for the study. Consistent with the previous rounds of the NELS, the fourth follow-up weights were modified to compensate for nonresponse and to provide point estimates of population totals that were consistent across rounds. The following discussion distinguishes between design or sampling weights and adjusted or analysis weights.

[44] Sampling weights are sometimes described as the inverses of the selection probabilities assigned to the units comprising the inferential population. Actually, the sampling weights are functions of these probabilities and also of the sample size and allocation. The term "frequencies" is intended to reflect this fact.

Table 3.8.3.2.—Summary Statistics for NELS:88 third follow-up weights: 1994

Statistic	F3QWT	F3PNLWT	F3F1PNWT	F3F2PNWT	F3CXTWT	F3PAQWT
Mean	214.67	226.25	226.45	218.21	227.69	251.57
Variance	55,899.72	61,822.48	60,950.67	57,695.52	130,221.50	85,368.41
Standard deviation	236.43	248.64	246.88	240.20	360.86	292.18
Coefficient of variation (×100)	110.14	109.90	109.02	110.08	158.49	116.14
Minimum	7.96	11.27	10.93	9.34	16.48	8.27
Maximum	6,135.13	7,549.94	7,521.50	7,118.84	12,444.78	8,358.50
Skewness	7.65	10.94	9.34	8.92	16.60	8.59
Kurtosis	108.61	211.61	163.12	147.95	428.73	142.01
Sum	3,201,743	2,968,426	3,160,792	3,201,743	2,677,913	3,197,396
Number of cases	14,915	13,120	13,958	14,673	11,761	12,710

Table 3.8.3.2.—Summary Statistics for NELS:88 third follow-up weights: 1994—continued

Statistic	F3TRSCWT	F3QWTG10	F3QWTG8	F3QWTG12	F3QWT92G
Mean	252.28	214.06	208.98	206.66	202.48
Variance	193,899.00	55,531	48,003.99	43,861.64	39,828
Standard deviation	440.34	235.65	219.10	209.43	199.57
Coefficient of variation (×100)	174.55	110.08	104.84	101.34	98.56
Minimum	7.20	7.96	7.96	16.23	7.96
Maximum	12,940	6,135.13	4,907.83	4,907.83	4,907.83
Skewness	10.78	7.69	6.92	7.23	7.37
Kurtosis	185.95	109.82	86.00	97.27	103.00
Sum	3,155,673	3,063,693	2,829,380	2,572,268	2,356,268
Number of cases	12,509	14,312	13,539	12,447	11,637

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1994.

3.8.4.1 Definition of Fourth Follow-up Weights

A total of nine different fourth follow-up analysis weights were computed for inclusion with the data files. The set of weights differentiates between cross-sectional and longitudinal (or panel) analyses for each of four data components, namely the basic questionnaire, contextual (i.e., teacher and administrator), parent, and high school transcript data components. The questionnaire weights were used with the data collected from the sample subjects themselves. The contextual data component consisted of school-level descriptive information about the teaching staff, policies, offerings, and school and student body characteristics. These data were collected from school administrators and teachers in the base-year and first two second follow-up rounds. Parent data, also collected in the base year and second follow-up, provide information about parental aspirations for their children, willingness and ability to commit resources for their children's education, home educational support system, and other family characteristics relating to educational achievement. The high school transcript component was added in the second follow-up and describes high school courses, grades, attendance, and postsecondary activities and achievement. The context, parent, and transcript components were used with the sample subject's questionnaire data to estimate parameters that describe the population of subjects. (The additional weights required for use with the postsecondary education transcript study data will be included with that data file. They are not described here.)

The variable names used to identify the nine weights follow the conventions adopted in earlier NELS data collection waves. A brief description of these weights follows:

F4QWT This is the fourth follow-up questionnaire weight. It applies to all fourth follow-up respondents. While the weight generalizes to no meaningful analysis population, when used in conjunction with appropriate population definitions, it can be used to estimate parameters that describe the populations of spring 1988 8th-graders in the year, 2000, regardless of 1988 and 1990 eligibility; who were eligible to complete questionnaires in 1992, 1994, and 2000, regardless of 1988 and 1990 eligibility; Spring 1990 10th-graders who were eligible to complete questionnaires in 1992, 1994, and 2000, regardless of 1988 and 1990 eligibility; and spring 1992 12th graders who were eligible to complete questionnaires in1992, 1994, and 2000.

F4BYPNWT This is the base-year and fourth follow-up panel weight and also the weight for base-year respondents who were also respondents to the fourth follow-up study. Response status in the other rounds was not considered. This weight applies to fourth follow-up respondents who were also respondents in the base-year survey (i.e., 1988). This weight is used to estimate longitudinal parameters that describe the population of spring 1988 8th-graders; it was used with the fourth follow-up descriptive report.

F4PNLWT This is the fourth follow-up complete panel weight, for respondents at all five NELS:88 data collection points. The weight applies to fourth follow-up respondents who were also respondents in each of the previous rounds (i.e., 1988, 1990, 1992, 1994). It is used to estimate longitudinal parameters that describe the population of spring 1988 8th-graders.

F4F1PNWT This is the first, second, third, and fourth follow-up panel weight. This weight applies to fourth follow-up respondents who were also respondents in the first, second, and third follow-up rounds (i.e., 1990, 1992, 1994). It can be used to

estimate longitudinal parameters that describe the population of spring 1990 10th-graders.

F4F2PNWT This is the second, third, and fourth follow-up panel weight, which applies to the 12th grade cohort. It applies to fourth follow-up respondents who were also respondents in the second and third follow-up rounds (i.e., 1992, 1994). It estimates longitudinal parameters that describe the population of spring 1992 12th graders.

F4CXTWT This weight is intended to be used with the 1992 school administrator and teacher data. It applies to NELS:88 2000 round respondents who were early graduates from or students in the spring of 1992 at the sampled second follow-up schools and who completed a 1992 questionnaire. (Teacher and school administrator data were collected from a subsample of the 1992 schools. This weight allows analysts to generate national statistics for students using the associated teacher and school administrator data despite the biases against small cluster sizes in sample selection.

F4PAQWT This weight is the parent data weight for fourth follow-up respondents with 1992 parent data. It can be used to estimate parameters that describe the population of fourth follow-up subjects, as subset with appropriate cohort flags.

F4TRSCWT This weight, the fourth follow-up high school transcript weight, is intended to be used with the fourth follow-up respondents who also have 1992 high school transcript data. The weight applies to year 2000 respondents whose second follow-up status was dropout, early graduate, or student in a sampled school and for whom transcripts were collected in 1992.

F4QWT92G This weight is equal to *F4QWT* for NELS:88 2000 wave respondents who received a high school diploma between September 1, 1991, and August 31, 1992, or respondents whose diploma receipt date is not known but who began their postsecondary education between June 1 and October 31, 1992. *F4QWT92G* is zero for all other respondents. *F4QWT92G* allows projections to the population of persons who received a high school diploma in the 1991-92 academic year.

Using the fourth follow-up weights in analysis. Data users are reminded that only the weights generated for the fourth follow-up study are included on the 1988-2000 public-use (N0P) and restricted-use (N0R) ECBs. (Apart from weights from earlier rounds, also not included are weights, or data, for the postsecondary transcript component of the fourth follow-up. The postsecondary transcript data will be released as a separate restricted-use ECB in 2002.).

Data users are urged to utilize the NELS:88 *universe variables* (F4UNIV1 and F4UNI2A through F4UNI2E) in planning their analyses. The universe variables can also be used to subset cases to the desired populations. The universe variables help determine the analytic group membership and round-by-round participation status of sample members. These variables track mode of entry into the sample, participation status by survey wave, in-school (student versus dropout) status, and in- or out-of- grade sequence (modal grade progression for cohort) status, and eligibility.

The general procedure for utilizing the weights in analysis involves subsetting cases through use of a flag or universe variable, then applying a WEIGHT statement to weight the cases to population totals. Below we further discuss use of flags and weights in both cross-sectional and longitudinal contexts.

Cross-sectional weights. An important preliminary question for any analysis is definition of the population of interest. The 1988-2000 sample of 12,144 individuals comprises three distinct populations and does not, by itself, constitute a meaningful analysis population.[45] The three meaningful analysis populations that it contains are, in 2000: spring-term 1988 8th-graders, spring-term 1990 sophomores, and spring-term 1992 high school seniors. Cross-sectional weights have been created to reflect questionnaire completion in 2000 as well as completion accompanied by various forms of supplemental data (1992 contextual data, 1992 parent survey data, high school transcripts), or graduation status (a special weight reflects questionnaire completion for the subset of year 2000 respondents who received a high school diploma in the 1991-92 academic year). These 2000 cross-sectional weights *(F4QWT, F4CXTWT, F4PAQWT, F4TRSCWT, F4QWT92G)* encompass all three analysis populations (that is, the NELS:88 representative grade cohorts). In turn, one can generalize about the status of respondents in 2000 in terms of any of these populations by invoking the appropriate cohort (the cohort flags are *G8COHORT*, *G10COHRT, G12COHRT*) flags.[46]

Longitudinal weights. Longitudinal (panel) weights are based on specific NELS:88 grade cohorts. Two 1988-2000 panel weights have been created for the NELS:88 1988 8th-grade cohort. *F4BYPNWT* takes account of participation only at the first and final data collections of the survey. The corresponding participation flag is *F4BYPNFL* (panel flag, member BY and F4).

On the other hand, *F4PNLWT* was created for 8th-grade cohort respondents who have data across all five NELS:88 data points. The corresponding flag is *F4PNLFL* (panel flag, member BY, F1, F2, F3 and F4). (There is no need to invoke the cohort flag since the panel flag already captures the 8th-grade cohort).

A single panel weight (*F4F1PNWT*) was created for the sophomore cohort for data from 1990 to 2000, for students with data in all four waves (the corresponding flag is *F4F1PNFL* [panel flag, member F1, F2, F3, and F4]). One panel weight was created for the NELS:88 senior cohort, for spring 1992 seniors with data in all three) relevant waves (1992, 1994, 2000). The corresponding flag is *F4F2PNFL* (panel flag, member F2, F3, and F4).

Weights have not been created for all conceivable NELS:88 analysis populations and objectives. However, in many instances, an existing weight can be substituted for a missing weight. For example, base-year parent data can be analyzed with the student weight. The entire base-year sample was eligible for the parent survey and a very high parent response rate was achieved. Under these circumstances, the student weight is a good approximation for a parent weight. On the other hand, in the second follow-up, only a subsample of students were eligible for the parent survey and the student weight therefore does not function well with the parent data. For this reason, special weights for including 1992 parent data have been generated in recent waves of NELS:88 (such as *F4PAQWT* in the fourth follow-up). Two considerations may help guide analysts in deciding whether one weight may be substituted for another. One consideration is the sum of the weights and how well it compares to true population totals. Another consideration is nonresponse bias, which should be analyzed to determine whether certain groups are disproportionately excluded by the weighting strategy.

[45] Without subsetting this group to reflect one of the three grade cohort memberships, the 12,144 weighted cases generalize to no pure grade cohort at all but rather to the population of the 8th grade students during the spring term of 1988 after it was supplemented with new students in 1990 and 1992 to make sophomore and senior cohorts.

[46] F4QWT92G can also be used to create a flag for use with the cohort flags (e.g., IF F4QWT92G > 0) to subset these cohorts. For example, using this indicator with G8COHORT would subset to spring 1988 8th-graders who graduated from high school "on schedule."

Generalizing to Populations with Fourth Follow-up Study Weights. By themselves, only the longitudinal NELS:88 fourth follow-up weights generalize to meaningful analysis populations. The population represented by the sum of the fourth follow-up questionnaire completion weight (F4QWT) is all sample members who completed a questionnaire in 2000, regardless of their cohort membership. This is not, in itself, a meaningful population to generalize to in analysis. For generalization using the 2000 cross-sectional weight, it is necessary to invoke a flag to identify one of the three analytically distinct grade cohorts contained in NELS:88. The three cohorts are (1) 8th-graders from the 1987-88 school year's spring term, (2) high school sophomores as of spring term,1990, and (3) spring-term 1992 high school seniors. In fact there is also a refined (or re-defined) version of the senior cohort weight: F4QWT92G applies to all sample members who received a high school diploma in the 1991-92 school year. The sum of the 2000 weights for each of these cohorts differs slightly from sums from earlier rounds. This is because population projections for 2000 take account of two forms of sample attrition: mortality, and expatriation, both of which reduce the size of the population to which generalization is made. (The fourth follow-up interview also collected information on high school graduation date that may have been missing from earlier NELS:88 waves.)

Longitudinal weights for 2000 can be used to generate population estimates on NELS:88 questionnaire variables for the analytically distinct cohort in which each is based. These longitudinal weights (also called panel weights) include contextual weights related to the cohort members' parents, teachers, school administrators, and high school transcripts. Table 3.8.4.1 summarizes the NELS:88 fourth follow-up weights and highlights the populations to which analysts may generalize. These descriptions are intentionally broad and do not include every possible sample exclusion or condition. Thus, the reader is referred to the additional discussion on the NELS:88 sampling and weighting activities elsewhere in this report (e.g., sections 3.2 through 3.8). Data users are cautioned that flags must be used to establish meaningful analysis populations when using the NELS:88 weights (see the account of F4QWT and the three NELS:88 cohort flags above for an example).

3.8.4.2 Overview of the Weight Calculations

The overriding principle in deriving analysis weights for NELS:88/2000 was to keep them as consistent as possible with those from the previous waves of data collection. When feasible, procedures and steps used in the previous follow-ups were adopted. However, since the fourth follow-up study was anticipated to be the last round of NELS:88 data collection, the weighting process deviated from the approach used in the third follow-up in several aspects. Notably, the eligibility criteria for sample members were altered. For example, several new categories (e.g., institutionalized/incarcerated) were created and these sample members were classified as ineligibles. Sample members in these categories would have been considered eligible in the previous waves of data collection. As a result, this could cause a slight change of the target population.

Another major change in the approaches used to adjust design weights for nonresponse and to complete post-stratification raking was also implemented in the fourth follow-up study. Sampling statisticians employed a Generalized Exponential Model for these activities. This innovative approach has a number of advantages over traditional methods based on weighting classes. (as discussed below.) Target totals used for post-stratification adjustment of the design weights were based on the analysis weights of the third follow-up.[47]

[47]It is important to note that the third follow-up expanded weight, which was required for creating the 4th Follow-up expanded weight, was not available during weighting for NELS:88/2000. Sampling statisticians recreated this weight from component parts from the third and second follow-up studies. Despite this "work-around," we do not believe this approach will have a significant impact on the final weights—there was no substantial change between the third and the fourth follow-up target populations estimates.

Table 3.8.4.1.—Descriptions and target populations for the NELS:88 fourth follow-up study analysis weights: 2000

Weight	Description
F4QWT	*Fourth Follow-up Study Questionnaire Weight.* This weight applies to all fourth follow-up respondents. Importantly, it must be used with appropriate sample flags (see page 86) to subset the respondent to a meaningful population. When so paired, this weight can be used by analysts to generalize to the following populations: ■ spring 1988 8th-graders who were eligible to complete questionnaires in 1992, 1994, and 2000, regardless of 1988 and 1990 eligibility, or in other words, this is approximately the population of students who were 8th-graders in spring 1988; ■ spring 1990 10th-graders who were eligible to complete questionnaires in 1992, 1994, and 2000, regardless of 1990 eligibility, or approximately the population of students who were 10th-graders in spring of 1990; and ■ spring 1992 12th-graders who were eligible to complete questionnaires in the second, third, and fourth follow-up studies. This approximates the population of 1992 spring term seniors.
F4QWT92G	*Fourth Follow-up 1992 High School Graduate Weight.* This weight is equal to F4QWT for NELS:88 2000 wave respondents who received a high school diploma between September 1, 1991 and August 31, 1992 or respondents whose diploma receipt date was not known but who began their postsecondary education between June 1 and October 31, 1992. F4QWT92G is 0 for all other respondents. In other words, F4QWT92G can be used to generalize to the population of persons who received high school diplomas during the 1991-1992 academic year.
F4PNLWT	*Fourth Follow-up Study Panel Weight.* This panel weight applies to sample members who completed questionnaires in all five rounds of NELS:88. In longitudinal analyses, F4PNLWT generalizes to the population of eligible spring 1988 8th-graders. Note that for all panel weights, flags should be used to establish populations.
F4BYPNWT	*Base-Year and Fourth Follow-up Panel Weight.* This weight was developed for sample members who were NELS:88 respondents in the first and last data collection waves for the study, regardless of their response in other waves. It allows projections in longitudinal analyses to the population of eligible spring 1988 8th-graders. Unlike the fourth follow-up panel weight (F4PNLWT), F4BYPNWT is based on only the end points for NELS:88 and, thus, is based on a larger sample size than the other panel weight, which required responses at each of the five NELS:88 waves.
F4F1PNWT	*Sophomore Cohort Panel Weight.* Based on the responses to the first through the fourth follow-up studies without regard to base-year respondent status, this weight can be used in longitudinal analyses to generalize to 10th-graders during the spring term of 1990 (the sophomore cohort).
F4F2PNWT	*Senior Cohort Panel Weight.* This weight is also known as the fourth to second follow-up panel weight. It applies to fourth follow-up respondents who were also respondents in 1992 and 1994. It allows projections in longitudinal analyses to the population of spring 1992 12th graders.
F4CXTWT	*Fourth Follow-up Study Contextual Weight*. This contextual weight is designed for use with NELS:88/2000 respondents, who also have teacher and school administrator data. Overall, this generalizes to the population of students who were 12th graders in spring 1992.
F4PAQWT	*Fourth Follow-up Study Parent Weight.* This weight is the parent data weight for fourth follow-up respondents with 1992 parent data. It was created from the responses of spring-term 1992 12th-graders who were eligible to complete questionnaires during the second, third and fourth follow-up studies, and for whom parent data was collected in 1992. This weight can be used to approximate the population of students who were 12th graders in spring 1992.
F4TRSCWT	*Fourth Follow-up High School Transcript Weight.* This weight applies to respondents in 2000 who were students, dropouts or early graduates in 1992, who have 1992 high school transcript data, and who were eligible to complete questionnaires in 1992 and 1994. (Use flags to subset to analysis groups.)

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88, 20000.

This section provides an overview of the calculation of the fourth follow-up weights, including the calculation of the design weights and their adjustment to compensate for unit nonresponse and to provide point estimates of population totals consistent across rounds. The process involved three steps described in the following sections:

Step 1. Calculate the fourth follow-up design weights. In general, the design weights are computed as the inverses of the expected selection frequencies assigned to each of the units comprising the inferential population. That is,

$$w_g = \pi_g^{-1}$$

where

π_g denotes the expected frequency with which the g^{th} unit appears in conceptually repeated samples selected given the sampling design, and,

w_g denotes the corresponding design weight.

One can identify the longitudinal observations obtained by the g^{th} unit by the subscript

$$\ell = 1,2,3,...$$

Then, for each follow-up round $(\ell = 2,...,5)$ the design weight becomes:

$$w_{\ell,g} = \left(\pi_{\ell-1,g}\pi_{\ell.g}\right)^{-1} = \frac{w_{\ell-1,g}}{\pi_{\ell,g}}.$$

Hence, the design weight calculations for the fourth follow-up began with the design weights for the third follow-up. For the weights unaffected by school sampling (questionnaire weights), the design weights were equal to the third follow-up design weights divided by the expected frequencies with which each of the third follow-up sample cases appeared in the fourth follow-up subsample. Since the fourth follow-up subsampling involved a second phase nonresponse sampling, these were the product of the expected frequencies with which sample members appeared in the fourth follow-up phase one subsample and the expected frequencies with which the phase one sampled members appeared in the further phase two subsample. For the weights requiring the association of a sample member with a sampled school or a sampled parent (transcript weight, contextual weight and parent weight), the design weights were obtained by further dividing the above design weights by the expected frequencies with which a school or a parent appeared in the second follow-up sample.

Step 2. Adjust for fourth follow-up unit nonresponse. The design weights were modified to compensate for nonresponse. The modifications took the form of adjustment factors, denoted by a'_g, that multiplied the design weights to produce the analysis weights. That is, the nonresponse-adjusted weights were:

$$w'_g = a'_g w_g.$$

The value of this adjustment factor varied depending on the status of a sample member during data collection as well as race, gender, and other characteristics.

Step 3. Multidimensional post-stratification raking. The nonresponse adjusted weights were typically further adjusted by making use of auxiliary information. The benefits of doing so are well recognized and include reductions in variance estimates and bias due to under/over coverage and consistent estimates from different sources or data collection waves (Deming and Stephan 1940, Zieschang 1990, Deville and Sarndal 1992). One shortcoming associated with this adjustment is that unlike the estimates based on design weights, the estimates based on adjusted weights are not, in general, unbiased. The raking adjustment of the weights had been done in the previous rounds primarily out of the concern that spurious changes in demographics or dropout rates might be introduced as a result of subsampling and exclusion of ineligible sample members across waves. In addition, the weights could become extreme in some small domains due to the adjustment for nonresponse and could inflate the variance estimate if not adjusted.

For the fourth follow-up, the convention that the total sums and marginal distributions of the weights on some important dimensions such as race, gender, dropout status and school characteristics equal those of the previous round excluding sample members ineligible for current round of data collection was adopted. The targeted sums of weight were developed separately for each of the fourth follow-up weights based on the corresponding third follow-up analysis weight. If we denote the post-stratification adjustment factors by a''_g, the final weights take the form:

$$w''_g = a''_g w'_g = a''_g a'_g w_g .$$

In previous rounds of NELS:88, a classification scheme was formed first, on which the adjustment of the design weights for nonresponse and post-stratification raking was done. Under this approach, for example, classes were defined in such a way that nonrespondents in the group tended to resemble the groups' respondents. The adjustment factors were computed as the ratios of the sum of the design weights over the units in a class divided by the sum of the design weights over the respondents in the same class. That is, the adjustment factors:

$$a'_g = \frac{\sum_{u_g \in S} w_g}{\sum_{u_g \in S} \delta_{r,g} w_g}$$

were computed for all values of the *g*-subscript contained in each of the classes. In this expression, $u_g \in S$ denotes units in the sample. The indicator variables:

$\delta_{r,g}$ = 1, if the g^{th} unit belongs to the respondent set *R*, and
= 0, otherwise.

In the fourth round, a model-based approach was taken by using the Generalized Exponential Model (GEM) (see e.g., Folsom and Singh 2001). Although methods based on weighting class can be used to perform the adjustment, and were used in earlier NELS:88 waves for student nonresponse adjustments,[48] the GEM-based approach is more flexible in that a targeted distribution can be easily

[48] In the NELS:88 base year, statisticians used logistic regressions of response propensity to adjust for *school* nonresponse. The fourth follow-up in 2000 employed the GEM model—logistic regression is a special case of the GEM—to adjust weights for *sample member* nonresponse. The GEM model gives more control over the adjustment factors used during the weight adjustment by automatically maintaining their size within a pre-specified range. This

imposed by including an explanatory variable in the model. This method is also preferred under circumstances when certain domains are collapsed because they are empty or contain only a few respondents—a situation common in NELS:88 after multiple data collection waves and subsample implementations. Under these circumstances, the weighting class adjustments fail to preserve the full sample weighted distributions. The modeling approach, on the other hand, preserves the full sample weighted distribution of the explanatory variables for all of the main and interaction effects that are included in the model.

GEM provides a unified approach to handle nonresponse adjustment and post-stratification adjustment by seeking minimum change in design weight while satisfying a set of constraints imposed on the final weight. Given the model, the adjustment factors are given by:

$$a'_g = 1 + \exp\left(X_g \hat{\beta}\right).$$

In this expression, X_g is a vector of characteristics observed for all of the units in the sample (i.e., explanatory variables) used to predict the observed response pattern. The estimated vector of regression coefficients, β, is determined to satisfy the set of constraints:

$$\sum_{ug \in S} w_g (1 + \exp(X_g \hat{\beta})) \delta_{r,g} X_g^T = \sum_{ug \in S} w_g X_g^T,$$

where X_g^T denotes the transpose of X_g. That is, the values β are chosen such that the sum of the adjusted weights over the respondent set is equal to the full sample weighted distribution over the variables included in the model.

The model and weighting class approaches are equivalent when the set of explanatory variables and the variables used to define the weighting classes are congruent. However, the model approach is preferred over weighting class adjustments under circumstances when certain of the classes are collapsed because they are empty or contain only a few respondents. Under these circumstances, the weighting class adjustments fail to preserve the weighted distributions of the full sample. The model approach, on the other hand, preserves the full-sample weighted distribution of the explanatory variables for all of the main and interaction effects that are included in the model.

Post-stratification adjustments were similarly computed.

3.8.4.3 Design Weight Calculations

Questionnaire design weight (F4QDWT). This weight was used to calculate the student questionnaire weight. It equals the third follow-up design weights divided by the expected selection frequencies.

In each of the two phases of sampling the fourth follow-up subsample was selected with probability proportional to size from within strata[49] that identified the sample member's response history. The size measures used in the procedure were the third follow-up weights. The size measures were first

control helps to minimize the unequal weighting effects of the adjustments. Logistic regression, on the other hand, groups the estimated response propensity into categories, and gives less control over the size of the adjustment factors (and less control over the unequal weighting effects).

[49] The design strata are, of course, post-strata, and the phase 1 and phase 2 fourth follow-up subsamples are applications of double-sampling for nonresponse. However, to avoid confusion with the post-stratification adjustment of the weights, the discussion refers simply to *strata*.

scaled so that their sum was forced to equal the sum of the third follow-up question-naire weights, *F3QWT*. The expected selection frequencies for the phase one and phase two subsamples were computed using these adjusted size-measures.

The subsamples were selected using a minimum replacement procedure under which the realized sampling frequencies differed from the expected frequencies by less than one (Chromy 1981). Units associated with very large size measures may, as a consequence, be "selected" more than once, depending on the stratum-level sample allocations. Otherwise the procedure was equivalent to without replacement sampling. Given the disparity in the size measures (i.e., design weights) that had arisen because of the cumulative subsampling in previous rounds, the minimum replacement randomization procedure has the advantage of providing strict probability proportional to size assignments to the units without the necessity of special treatment for self-representing units (i.e., units with size measures sufficiently large that they must appear at least once in any sample selected given the design).

If we denote the size measure associated with the i^{th} individual in the h^{th} stratum by $S_{h,i}$, then the expected number of times the i^{th} individual is selected given the stratum-level sample allocation is:

$$E_{h,i} = n_h \frac{S_{h,i}}{S_h},$$

where

$n_h = \sum_{i=1}^{N_h} E_{h,i}$ is the stratum-level sample allocation,

$S_h = \sum_{i=1}^{N_h} S_{h,i}$ is the total size of the stratum, and

N_h is the number of population units in the stratum.

The expected frequency with which the i^{th} individual appears in samples of size n_h was computed as

$$\pi_{h,i} = \frac{n_h S_{h,i}}{n_{h,i} S_h},$$

where $n_{h,i}$ is the observed number of times the i^{th} individual is selected.

Contextual design weight (F4CXDWT). In 1992, contextual (or teacher/school administrator) data were collected from a subsample of schools. Hence, the contextual design weight for a fourth follow-up subject is the fourth follow-up questionnaire design weight (F4QDWT) divided by the expected frequency with which the 1992 school attended by the sample subject was selected into the contextual data component subsample from that year.[50] If $\pi_{k|i}$ denotes the expected frequency with which the k^{th} school attended by the i^{th} 1992 student was selected into the contextual data subsample, then the 1992 contextual data design weight is similarly obtained by dividing the first follow-up design weight, F1DWT, by π_k, producing:

[50] The 1992 school subsampling selection frequencies developed during the third follow-up study were not archived, and fourth follow-up study statisticians were required to recalculate these figures.

$$F2CXDWT_i = F1DWT_i / \pi_{k|i} .$$

This leads to the fourth follow-up contextual data design weight being computed as:

$$F4CXTDWT_i = F4QDWT_i \times F2CXTDWT_i / F1DWT_i ,$$

for the i^{th} fourth follow-up sample subject. However, the use of $F1DWT_i$ would exclude the second follow-up freshened sample members. Since $F2DWT_i$ was equal to $F1DWT_i$ for a first follow-up sample member and was equal to $F1DWT_i$ for a first follow-up sample member linked to a second follow-up freshened student, $F2DWT_i$ was used in the place of $F1DWT_I$.

Parent design weight (F4PADWT). Collection of parent data was initially conducted for all of the 1992 sample students. However, toward the end of the data collection period for the second follow-up study in 1992, the students for whom parent data had not yet been received were subsampled and parents in the subsample were made the object of an intensive data collection effort. Strata were constructed to identify Hispanics, Asians, high SES Whites, and "other" race-ethnic group members and subsampling was undertaken within these strata.

The stratum-level subsampling rates that were used during this subsample in 1992 were carefully documented, however, the information required to identify the specific individuals belonging to the initial and nonresponse second follow-up subsamples was not maintained in the available documentation. The documentation further reported that subsampling occurred when 70 percent of the parent data collection had been completed. As a reasonable approximation, stratum-level average frequencies were computed as:

$$\pi_i = (1.0)(0.70) + \theta_k (0.30)$$

for all values of the *i*-subscript contained in the h^{th} stratum. The value of θ_k is the documented nonresponse subsampling rate for stratum *h*. The effect is, of course, to apply the stratum-level average subsampling rates over all individuals in the stratum (i.e., over Hispanics, Asians, high SES Whites, and "others"). For the affected strata, the parent data design weights were then computed as:

$$F4PADWT_i = F4QDWT_i / \pi_i .$$

High school transcript design weight (F4TRDWT). The design weights for the NELS:88 high school transcript data were similarly computed. Using the same notation, the transcript data design weights were computed as:

$$F4TRDWT_i = F4QDWT_i \times F2TRDWT_i / F2DWT_i .$$

Although the same subsample of schools was used for each of the transcript and contextual data components, some differences between the 1992 transcript and contextual design weights were noted. Thus, the *F4TRDWT* and *F4CXTDWT* values are not identical.

3.8.4.4 Nonresponse Adjustments

Table 3.8.4.4 presents the explanatory variables used for computing the nonresponse adjustment factors. Data from the fourth follow-up round were used for student-level questionnaire weights, but contextual and transcript weights used school characteristics taken from the second follow-up wave. Missing and unknown variable values were treated as a separate category when the number of sampled cases in these categories was sufficiently large. The variables "race/ethnicity" and "ever dropped out at second follow-up" were combined in the fourth follow-up in order to highlight the importance of dropout status in the study. A separate category was created for dropouts at second follow-up, regardless of their race/ethnicity.

The frequency distributions of the numbers of respondents in each of the categories in table 3.8.4.4 were examined to identify categories with few respondents. Some categories were collapsed with others so that each category contained a reasonable proportion of respondents. The collapsed categories tended to be the same across data components (e.g., base-year school type variables were recoded to public, Catholic, and other private). The estimated adjustment factors ranged from 1 to 2.5, except for a small number of sample members who were classified as nonrespondents or poor respondents in the third follow-up data collection. Separate adjustments for these members were done and the adjustment factors were higher due to the low response rate among these members.

Table 3.8.4.4.—Factors used to adjust for nonresponse in the NELS:88 fourth follow-up study: 2000

Category	Nonresponse adjustment factor
Race-ethnicity	White, non-Hispanic Black, non-Hispanic Hispanic American Indian/Alaska Native
Ever dropped out at second follow-up	Yes No
Gender	Male Female
Base-year or second follow-up* school type	Public Catholic Other religion Nonreligious
Base-year or second follow-up school urbanity	Urban Suburban Rural
Base-year or second follow-up school region	Northeast Midwest South West

*Base-year variables were used for the questionnaire component, and second follow-up variables were used for other weight components.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

3.8.4.5 Post-stratification Adjustments

Post-stratification raking for the fourth follow-up weights included the variables used for nonresponse adjustment (see table 3.8.4.4), as well as some additional variables. These variables indicated the status of a sample member in the previous rounds of the survey (see table 3.8.4.5). The consistent distributions of the weights by these variables are important to ensure consistent point estimates across rounds of the study. As with nonresponse variables, categories with few respondents were combined with others. The estimated adjustment factors ranged from 0.5 to 2.5.

Table 3.8.4.5.—Factors used to post-stratify NELS:88 fourth follow-up study weights: 2000

Category	Post-stratification adjustment factor
Base-year response status	Dropout Eligible Ineligible
First follow-up response status	Dropout In school, in 10th grade In school, not in 10th grade Ineligible, in school, in 10th grade Ineligible, in school, not in 10th grade Ineligible, not in school Deceased or out of country Freshened, dropout Freshened, in school Freshened, ineligible
Second follow-up response status	Dropout In school, in 12th grade In school, not in 12th grade Ineligible, in school, in 12th grade Ineligible, in school, not in 12th grade Ineligible, not in school Deceased or out of country Freshened, dropout Freshened, in school Freshened, ineligible
Third follow-up response status	Eligible, received high school diploma Eligible, received graduation equivalency diploma or certificate of attendance Eligible, currently pursuing diploma or equivalent Eligible, did not receive diploma or equivalent and not pursuing one
Eligibility for inclusion in 8th–10th grade dropout rate	Ineligible Eligible, dropout in the first follow-up Eligible, not a dropout in the first follow-up Eligible for inclusion in the calculation of 10th to 12th grade dropout rate Eligible, dropout in the second follow-up Eligible, not a dropout in the second follow-up

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

3.8.4.6 Summary of Weight Distributions

Summaries of the cross-sectional and panel weight for the fourth follow-up respondents are included in table 3.8.4.6-A. Table 3.8.4.6-B provides the distributions for the fourth follow-up study's contextual, parent, and high school transcript weights. The summary statistics shown in the tables follow those used in earlier NELS reports. Compared to the final weights from the previous rounds, the means and variances of the fourth follow-up weights were larger—as expected as a result of multiple subsampling activities that took place during the round. A notable change is the large increase in the maximum value of the fourth follow-up weights. For example, the largest F4QWT was 20,899, while the corresponding F1QWT, F2QWT and F3QWT had a maximum value of 6,697, 6,670, and 6,135. To a large extent, this increase was due to the second phase nonresponse subsample in the fourth follow-up study. (The fact that only 219 sample members were selected during the subsample and that these sample members were selected from four domains,[51] the overall impact of these very large weights on analysis should be small. However, data users should be cautious when analyses are restricted to the domains with which these sample members are associated.)

Table 3.8.4.6-A.—Summary of questionnaire and panel weight distributions for the NELS:88 fourth follow-up study: 1988–2000

Item	F4QWT	F4QWT92G	F4BYPNWT	F4PNLWT	F4F1PNWT	F4F2PNWT
Mean	259.27	239.03	257.20	269.84	272.25	264.28
Variance	162,623.40	127,593.79	262,329.59	107,616.01	188,396.23	164,294.53
Standard deviation	403.27	357.20	512.18	328.05	434.05	405.33
Coefficient of variation ×100	155.54	149.44	199.14	121.57	159.43	153.37
Minimum value	8.71	8.71	9.81	11.11	7.86	7.29
Maximum value	20,898.71	20,898.71	25,938.81	10,370.88	11,290.03	9,822.85
Skewness coefficient	16.68	23.76	19.24	11.09	9.30	8.27
Kurtosis coefficient	629.81	1,168.76	708.27	213.50	139.62	107.56
Sum of weights	3,148,607.79	2,392,699.45	2,927,935.48	2,921,547.08	3,109,594.75	3,148,607.80
Number of cases	12,144	10,010	11,384	10,827	11,422	11,914

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-2000.

[51]These domains included dropouts, American Indians and Alaska Natives, Black non-Hispanics, and Hispanics.

Table 3.8.4.6-B.—Summary of weight distributions for contextual, parent, and high school transcript weights for the NELS:88 fourth follow-up study: 1988-2000

Item	Contextual F4CXTWT	Parent F4PAQWT	Transcript F4TRSCWT
Mean	268.11	297.11	301.74
Variance	238,934.96	163,960.52	301,606.94
Standard deviation	488.81	404.92	549.19
Coefficient of variation ×100	182.31	136.29	182.01
Minimum value	18.58	10.29	7.39
Maximum value	14,956.56	18,926.14	18,197.05
Skewness coefficient	13.04	15.62	13.16
Kurtosis coefficient	249.90	515.29	274.26
Sum of weights	2,646,830.00	3,150,858.35	3,110,890.68
Number of cases	9,872	10,605	10,310

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-2000.

3.9 Standard Errors and Design Effects

This section examines the calculation of standard errors as a measure of sampling variability in the NELS:88 survey. The standard error is an estimate of the expected difference between a statistic from a particular sample and the corresponding population value. Variance estimation procedures and results are discussed below with advice to analysts for the use of NELS:88 design effect information.

3.9.1 Introduction: Estimating Standard Errors and Design Effects

Survey Standard Errors. Because the NELS:88 sample design involved stratification, disproportionate sampling of certain strata, and clustered (i.e., multistage) probability sampling, the resulting estimates are more variable than they would have been had they been based on data from a simple random sample of the same size.

The calculation of exact standard errors for survey estimates can be difficult and expensive. Popular statistical analysis packages such as SAS and SPSS typically calculate standard errors without accounting for the complex sample designs associated with longitudinal studies like NELS:88. This may have important consequences with significance testing, which will be based on statistical comparisons made with artificially small standard errors and upwardly biased significance levels. Several procedures are available for calculating precise estimates of sampling errors for complex samples. These procedures, most commonly Taylor series approximations, balanced repeated replication (BRR), and jackknife repeated replication (JRR) approaches, produce generally comparable results. (See Frankel 1971 for an overview of these approaches.) Consequently, it is largely a matter of convenience which approach is taken. For NELS:88, the Taylor series procedure has typically been used to calculate the standard errors. The explanation below outlines the concepts that underpin the procedure used to calculate standard errors.

In a simple random sample, the mean is estimated as:

$$\hat{x} = \sum x_i \Big/ n .$$

Only the numerator (i.e., the sample total) is subject to sampling error; the denominator (the sample size) is fixed. In more complex designs, such as the NELS:88 design, the mean is estimated as a ratio of the estimates; for NELS:88, this ratio can be expressed as

$$r = \sum y_{hij} \Big/ \sum x_{hij} ,$$

in which y_{hij} is the weighted value for student *j* from school *i* in stratum *h,* and x_{hij} is the weight for that student. The numerator in the previous equation is an estimate of the relevant population total; the denominator is an estimate of the population size. Both estimates are subject to sampling error.

Kish (1965) has shown that the variance of *r* (as defined in the equation above) is:

$$E(r-R)^2 = E\left[\frac{dy - Rdx}{X} * \frac{1}{1+dx/X}\right]^2 ,$$

in which

$E(r-R)^2$ = the expected value of the squared difference between the population ratio, *R,* and the sample estimate, *r*;

dy = the difference between the sample estimate of the population total, *y*, and the population total, *Y*;

X = the population size;

x = the difference between the sample estimate of the population size, *x,* and the actual population size, *X.*

If the term involving the relative error in the estimate of the population size (*dx/X*) is ignored, this equation reduces to

$$E(r-R)^2 = 1 \Big/ X^2 \left[Var(y) + Var(x) - 2RCov(xy)\right].$$

In this equation, Var(*y*) and Var(*x*) refer to the variance of *y* and *x,* and Cov(*xy*) refers to their covariance. All of these terms can be estimated from sample data (i.e., *r* would replace *R, x* would replace *X,* and so on).

Estimates of variance terms are based on the variation of the individual school means around the estimated stratum mean. Various rationales have been offered for the use of this equation as an approximation to the preceding equation. One line of argument is based on a standard mathematical tool called Taylor series approximation, the approach that was used in NELS:88.

Design Effects. The impact of departures from simple random sampling on the precision of sample estimates is often measured by the design effect (designated as *DEFF*). For any statistical estimator (for example, a mean or a proportion), the design effect is the ratio of the estimate of the variance of a statistic derived from consideration of the sample design to that obtained from the formula for simple random samples (i.e., var(y)/n). The square root of the design effect (also called the root design effect, and designated as *DEFT*) is also useful. The following formulas define the DEFFs and DEFT for this section:

$$DEFF = \frac{(Design\ SE)^2}{(SRS\ SE)^2}$$

and

$$DEFT = \frac{Design\ SE}{SRS\ SE},$$

where *DESIGN SE* designates the standard error of an estimate calculated by taking into account the complex nature of the survey design, and *SRS SE* designates the standard error of the same estimate calculated as if the survey design was a simple random sample.

3.9.2 Calculation of Design Effects

SUDAAN, a data analysis program designed specifically to correctly estimate standard errors for data collected with complex surveys, was used to calculate design effects for the NELS:88 second and fourth follow-up studies (see, e.g., Shah, Barnwell, and Bieler 1997, for information on the program). For 1988 and 1990 data, the design effects were calculated by taking the ratio of the standard error adjusted for the sampling design, obtained from a standard Taylor series linearization program, divided by the weighted simple random sample standard error obtained from a statistical analysis program such as SAS. For the third follow-up study, standard error and design effects were also calculated using Taylor series approximations, this time available from the NELS:88/94 data analysis system (DAS).[52] The estimation procedures and approximations used by SUDAAN for calculating design effects differ both quantitatively and qualitatively from methods used in past rounds. Notably, in certain circumstances, large discrepancies can occur between SUDAAN-calculated design effects and those used in previous rounds.

These differences involve the SUDAAN program's estimation of the simple random sample standard error used in the denominator of the design effect. In its design effect calculation, SUDAAN uses an unconditional estimate of the simple random sample standard error based on the estimated proportion of subgroup respondents in the population. Design effects calculated for previous rounds of NELS, however, used a simple random sample standard error based on the proportion of the subgroup respondents in the sample (conditional estimate). The two standard error estimates are different because of oversampling and nonresponse. For example, if there were 3,000 Hispanics in a sample and Hispanics were oversampled at twice the rate of their proportion in the population, the conditional simple random sample standard error estimate for Hispanics would be based on an sample size of 3,000. For its unconditional estimate, however, SUDAAN would base the design effect on half of that sample size, an N of 1,500. Basing the denominator standard error on an N of 3,000, which is comparable to the way design

[52] The DAS is an NCES product that uses Taylor series approximations to develop tables of estimates and standard errors for complex sample designs.

effects were calculated in previous rounds of NELS, would give a larger design effect (i.e., a smaller simple random sample standard error) than basing it on the *N* of 1,500.

3.9.3 Selection of Items Used to Estimate Standard Errors and Design Effects

Selection of Base-Year Items. Standard errors and design effects were selected for 30 means and proportions based on the NELS:88 base-year student, parent, and school data. The 30 variables from the student questionnaire were selected to overlap as much as possible with those variables examined HS&B. The remaining variables from the student questionnaire and from the parent and school questionnaires were selected randomly from each topical section of the questionnaire. Standard errors and design effects were calculated for each statistic both for the sample as a whole and for selected subgroups. For both the student and parent analyses, the subgroups were based on the student's sex, race-ethnicity, school type (public, Catholic, and other private), and SES (lowest quartile, middle two quartiles, and highest quartile). For the school analysis, the subgroups were based on two levels of school type (public and combined private) and 8th grade enrollment (at or below the median and above the median).

Selection of Items in the 1990 and 1992 Follow-up Waves. Standard errors and design effects were also calculated for 30 means and proportions based on the NELS:88 first follow-up student and dropout data. The goal was to estimate standard errors/design effects for all respondents including dropouts, on the one hand, and separately for dropouts, on the other. Because of the lack of perfect overlap between questions on the student and dropout questionnaires, and because 25 percent of the dropout sample was administered an abbreviated questionnaire, it was necessary to select two sets of 30 items, one to represent questions asked of all respondents and one to represent questions asked of all dropouts.

To select questions for the standard errors/design effects analysis of all respondents, a number of criteria were used. The first criterion was whether a question appeared in the NELS:88 base-year or HS&B analyses of standard errors/design effects. This criterion resulted in the selection of ten questions, seven of which were used in both the NELS:88 base-year and HS&B standard error/design effects analysis and three of which were used only in the NELS:88 base-year analysis.

Policy relevance was the second criterion used for selecting questions. This criterion was used to ensure that variables important to analysts were represented. Using this criterion, four cognitive test scores, specifically the IRT (Item Response Theory)-estimated number-right scores for math, English, science, and social studies, were selected. Although several test score composites are available in the data file, the IRT-estimated number-right scores were chosen because they compensate for guessing and for omitted items. The IRT scores also have the virtue of being equated across the multilevel math and reading test forms.

The remaining 16 variables were selected randomly from the pool of remaining critical items.[53] The selection process used the following procedure. First, all critical items not selected by the first two criteria formed a pool of eligible items. This involved three types of items—binary items, multiple-category items, and continuous or quasi-continuous items. Each category of a multiple-category item was treated as a separate binary item. Second, all of the items (binary and continuous) were rescaled such that the lowest possible value was 0 and the highest possible value was 100. Finally, the rescaled items were sorted by the size of their means, and a systematic sample of 16 items was selected from the sorted list of items.

[53] A "critical" item is a member of the subset of questions that, if not answered, trigger an attempt to recontact the respondent to obtain the missing data.

For dropouts, the starting point for selecting the variables for standard error/design effect calculations was to use items that overlapped in the student and dropout questionnaires and that were already selected for the analysis of all respondents. There were 18 such items. The remaining items were selected randomly from the pool of critical items not already selected that were in both the full and abbreviated versions of the dropout questionnaire. A systematic sample of 12 items from this pool was obtained by the same transformation, ordering, and systematic sampling procedure used to select items for all students.

Standard errors and design effects were also calculated for 30 means and proportions based on the NELS:88 second follow-up student and dropout data. As in the first follow-up analysis, the goal was to estimate standard errors/design effects for all respondents including dropouts, and separately for dropouts. Second follow-up design effects were also calculated for the transcript and parent surveys, as well as the contextual sample (which was eligible for school administrator and math or science teacher reports).

Criteria similar to those used in the first follow-up were used to select questions for the second follow-up standard error/design effects analysis. The first criterion was whether a question had been used in the NELS:88 base-year and first follow-up or HS&B analyses of standard errors/design effects. This overlap resulted in the inclusion of 16 items. Additionally, it was important to maximize the overlap between questions that appeared in both the second follow-up student and dropout questionnaires. Nine of the remaining items selected appear in both second follow-up instruments. A total of five non-overlap items were selected from the student questionnaire to supplement those in common with the dropout questionnaire. Policy relevance was the second criterion for selecting items in the second follow-up. Using this criterion, four cognitive test scores were selected—the IRT-estimated number-right scores for mathematics, English, science, and social studies.

Selection of Items in the Third Follow-up Study. Standard errors and design effects were calculated for 30 means and proportions based on the NELS:88 third follow-up student and dropout data. As in the previous rounds, the goal was to estimate standard errors/design effects for all respondents, including dropouts. Criteria similar to those used in the second follow-up were used to select questions for the third follow-up standard error/design effects analysis. The first criterion was whether a question had been used in the NELS:88 analysis of standard errors/design effects in any of the previous rounds. This overlap resulted in the inclusion of five items. Additional items were then chosen if they appeared in the crosswalk of the other rounds. Sixteen of the remaining items selected appeared in one or more of the previous rounds. The remaining nine items were chosen at random from the third follow-up such that three items involved information about postsecondary education, three pertained to work activity, and three involved personal information about the respondent.

Selection of Items in the Fourth Follow-up Study. Procedures for calculating standard errors and design effects for the fourth follow-up study modeled the approaches used in the previous follow-up study. Thirty means and proportions based on the NELS:88/2000 data were used to estimate standard error/design effects for all respondents. The research team selected ten variables to represent the earlier rounds of the study: five from the second follow-up study in 1992 and five from the third follow-up study in 1994. The remaining items were picked randomly to represent each section of the NELS:88/2000 CATI/CAPI interview.

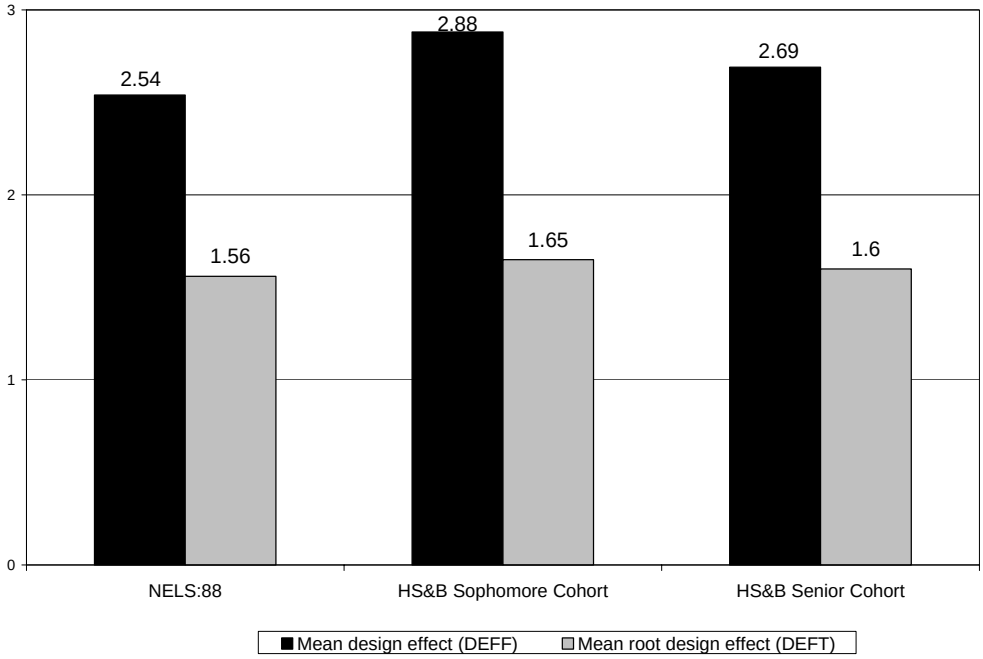
Standard errors and design effects were calculated for each of the items for the sample as a whole, including students and dropouts. The analyses were then repeated for the key respondent race-ethnicity and gender subgroups. Standard errors and design effects were calculated using the fourth follow-up respondents weighted by the full sample questionnaire weight (*F4QWT*). (The tables of results are included as tables D-1 through D-10 in appendix D.)

3.9.4 Variance Estimation Results: Standard Errors and Design Effects

Summary results appear in this chapter; item-level results for the final wave of NELS:88 may be found in appendix D. Further standard error tables at the item level, used to compute design effects for subgroups, can be found in other NELS:88 documents. Specifically, standard errors for individual items are presented by subgroup for 1988 in the respective base-year user's manuals and sample design report; for the 1990 round, see the *NELS:88 First FollowUp Final Technical Report (*NCES 94-632); for 1992, the *NELS:88 Base-Year through Second FollowUp Final Methodology Report* (NCES 98-06).

Base-Year Results. On the whole, the design effects for the NELS:88 base-year study indicate that the NELS:88 sample was more efficient than the HS&B sample.[54] For means and proportions based on student questionnaire data for all NELS:88 and HS&B students (see figure 3.9.4), the average design effect in the NELS:88 base year was 2.54; the comparable base-year figure was 2.88 for the HS&B sophomore cohort and 2.69 for the senior cohort.

Figure 3.9.4.—Full sample design effects for the NELS:88 base-year study and HS&B sophomore and senior cohorts: 1980 and 1988



SOURCE: U.S. Department of Education, National Center for Education Statistics, High School and Beyond (HS&B) and National Education Longitudinal Study of 1988 (NELS:88), 1980 and 1988.

[54] Standard errors and design effects for questions selected from the base-year student questionnaire are presented in section 3.3 of Ingels, Dowd, Baldridge, Stipe, Bartot and Frankel, 1994 (tables 3.3-1 and 3.3-2) (NCES 94-374).

Table 3.9.4-A gives the mean DEFFs and mean DEFTs for selected subgroups of base-year respondents. This table indicates that the difference in NELS and HS&B design effects was also apparent for subgroup estimates. The *High School and Beyond Sample Design Report* presents design effects for ten subgroups (Frankel et al. 1981). For eight of the ten subgroups, the NELS:88 base-year average design effects were smaller than those for both the HS&B sophomore and senior cohorts. The increased efficiency is especially marked for students attending Catholic schools. In NELS:88, the average design effect was 2.70; in HS&B, it was 3.60 for the sophomore cohort and 3.58 for the senior cohort.

These design effects in the NELS:88 base year may reflect the somewhat smaller cluster sizes used in the later surveys. The HS&B base-year sample design called for 36 sophomore and 36 senior selections from each school; the NELS:88 sample called for the selection of only 24 students (plus, on average, two oversampled Hispanics and Asians) from each school. Clustering tends to increase the variability of survey estimates because the observations within a cluster are similar and therefore add less information than independently selected observations.

First Follow-up Results. Standard errors and design effects were calculated for each of the 30 items selected for the first follow-up sample as a whole and for selected subgroups. The subgroups were based on the respondent's school status (student/dropout), sex, race-ethnicity, school type (public, Catholic, and other private), SES (lowest quartile, middle two quartiles, and highest quartile), and urbanicity (urban, suburban, and rural). Two sets of standard errors and design effects were calculated, one using all of the first follow-up respondents weighted by the full sample questionnaire weight, *F1QWT,* and the second using just the panel respondents weighted by *F1PNLWT*.

Table 3.9.4-B presents summary design effects for the first follow-up study's full respondent sample. (Individual item standard errors, design effects, and design effect summary statistics for dropouts are presented in Ingels et al. 1994) The sample sizes for the subgroup analyses were too small to estimate design effects for the dropouts.

Second Follow-up Student and Dropout Results. Standard errors and design effects were calculated for each of the items for the sample as a whole and for selected subgroups. The subgroups were based on the respondent's sex, race-ethnicity, school type (public, Catholic, and other private), SES (lowest quartile, middle two quartiles, and highest quartile), and urbanicity (urban, suburban, and rural). Three sets of standard errors and design effects were calculated, one using all of the second follow-up respondents weighted by the full sample questionnaire weight (*F2QWT);* the second using the panel respondents weighted by *F2PNLWT;* and the third using the first and second follow-up panel sample members weighted by *F2F1PNWT*. Table 3.9.4-C presents corresponding summary design effects for the full sample's subgroups. The individual item standard errors, DEFFs, and DEFTs for all respondents are presented along with summary statistics in Ingels et al. (1994) (NCES 94-374).

Table 3.9.4-A.—Mean design effects (DEFFs) and root design effects (DEFTs) for selected sampling strata for base-year student questionnaire data: 1988

Group	Mean DEFF	Mean DEFT
All students	2.54	1.56
Male[1]	1.98	1.39
Female	1.93	1.38
White and other[2]	2.25	1.48
Black	1.65	1.27
Hispanic	2.06	1.41
Asian/Pacific Islander	2.00	1.40
Public schools	2.27	1.48
Catholic schools	2.70	1.59
Other private schools	8.80	1.83
Low SES	1.58	1.25
Middle SES	1.66	1.28
High SES	1.84	1.34

[1]Sex categories are based on the composite sex variable.
[2]Race categories are based on the composite race variable.

NOTE: Each mean is based on 30 items, including four cognitive test score composites. Although this table does not reflect the rescaling of base-year cognitive test results in the second follow-up, the correlation between the cognitive test items before and after the escaling is 0.99.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

Table 3.9.4-B.—Mean design effects (DEFFs) and root design effects (DEFTs) for selected sampling strata for the NELS:88 first follow-up study's *full sample* of students and dropouts:[1] 1990

Group	Mean DEFF	Mean DEFT
Students	3.86	1.92
Dropouts	4.71	2.00
Male[2]	3.37	1.80
Female	3.45	1.81
White	3.05	1.71
Black	3.62	1.83
Hispanic	3.56	1.76
Asian/Pacific Islander	2.77	1.63
American Indian/AK Native	2.42	1.44
Public schools	3.23	1.76
Catholic schools	2.67	1.54
Other private schools	6.65	2.42
Low SES	2.84	1.65
Middle SES	3.09	1.72
High SES	3.48	1.80
Urban	3.48	1.85
Suburban	3.48	1.80
Rural	2.67	1.58

[1]This table is based on the original (1992-93) release of the first follow-up student file. The second follow-up (1994) release of the first follow-up student data contains a slightly different sample number than the original release. See Section 3.1.2 of the *NELS:88 Second Follow-Up: Student Component Data File User's Manual* for additional details about the sample numbers of the two releases.

[2]Sex categories are based on the composite sex variable.

NOTE: Each mean is based on 30 items, including four test score composites. Although this table does not reflect the rescaling of first follow-up NELS:88 achievement battery items in the second follow-up, the correlation between the test items before and after the rescaling is 0.99.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1990.

Table 3.9.4-C.—Mean design effects (DEFFs) and root design effects (DEFTs) for selected sampling strata for the NELS:88 first follow-up study's *panel sample*[1] of students and dropouts:[1] 1988-1990

Group	Mean DEFF	Mean DEFT
Students	3.80	1.91
Dropouts	4.71	2.00
Male[2]	3.46	1.82
Female	3.32	1.78
White	3.10	1.73
Black	3.80	1.87
Hispanic	2.64	1.59
Asian	2.76	1.61
American Indian/Alaska Native	2.07	1.36
Public schools	3.15	1.74
Catholic schools	2.62	1.51
Other private schools	6.53	2.39
Low SES	2.80	1.64
Middle SES	3.14	1.73
High SES	3.58	1.82
Urban	3.46	1.84
Suburban	3.41	1.79
Rural	2.63	1.57

[1]This table is based on the original (1992-93) release of the first follow-up student file. The second follow-up (1994) release of the first follow-up student data contains a slightly different sample number than the original release. See Section 3.1.2 of the *NELS:88 Second Follow-Up: Student Component Data File User's Manual* for additional details about the sample numbers of the two releases.

[2]Sex categories are based on the composite sex variable.

NOTE: Each mean is based on 30 items, including four test score composites. Although this table does not reflect the rescaling of first follow-up cognitive test items in the second follow-up, the correlation between the test items before and after the rescaling is 0.99.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-90.

3.9.5 NELS:88 Third Follow-up Standard Errors and Design Effects Results

Standard errors and design effects were calculated for each of the items for the sample as a whole, including students and dropouts. The analyses were then repeated for the 17 sampling subgroups.[55] Standard errors and design effects were calculated using the third follow-up respondents weighted by the full sample questionnaire design weight (F3QWT). The individual item standard errors, DEFFs, and DEFTs for all respondents and all subgroupings can be found, along with summary statistics in Haggerty, Dugoni, Reed, Cederlund, and Taylor 1996. Summary overall, dropout, and private school student results are also presented in the table 3.9.5 below.

Table 3.9.5.—Mean design effects (DEFFs) and root design effects (DEFTs) for selected sampling strata for the NELS:88 third follow-up study's full sample: 1994

Group	Mean DEFF	Mean DEFT
All	2.94	1.70
Dropouts	2.79	1.65
Private high school in 1990 or 1992	1.58	1.24
Private middle school in 1988	4.43	2.05

NOTE: Each mean is based on 30 items.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1994.

3.9.6 Estimated Fourth Follow-up Standard Errors and Design Effects

Standard errors, the square root of the variance as a measure of sampling variability, and the design effects and root design effects as measures of the impact of stratification, clustering, and unequal weighting effects were calculated for selected fourth follow-up questions. The parameters of the selected items were weighted by the fourth follow-up questionnaire weight, *F4QWT*. Standard errors and design effects were estimated using SUDAAN, and analyses were completed for all fourth follow-up respondents and for subgroups defined by respondents' gender, race-ethnicity, and dropout status. Strata containing only one primary sampling unit (PSU) as a result of subsampling were collapsed to their nearest strata. The results, along with some summary statistics of the estimated design effects, are presented in appendix D of this report (table D-1 through table D-10). Summary overall, dropout, and student results, along with results for selected racial ethnic and gender subgroups are presented in table 3.9.6.

3.9.7 Comparison of Standard Errors and Design Effects for all NELS:88 Waves

Table 3.9.7 provides summary statistics for design effects for all five waves of the study. Figure 3.9.7 provides a graphic display of the mean DEFFs and mean DEFTs for these waves.

[55] Four of the sampling subgroups were omitted from the design effect analysis because of insufficient sample size. These were (1) nonresponders, (2) 1990 freshened, (3) 1992 freshened, and (4) other.

Table 3.9.6.—Mean design effects (DEFFs) and root design effects (DEFTs) for selected sampling strata for the NELS:88 fourth follow-up study: 2000

Group	Mean DEFF	Mean DEFT
All respondents	3.93	1.95
Male	4.17	2.01
Female	2.93	1.70
Asian/Pacific Islander	3.09	1.74
Hispanic respondents	4.24	2.05
Black, non-Hispanic	5.92	2.36
White, non-Hispanic	2.71	1.63
American Indian and Alaska Native	3.31	1.75
Dropouts	3.02	1.72
Non-dropouts	3.75	1.91

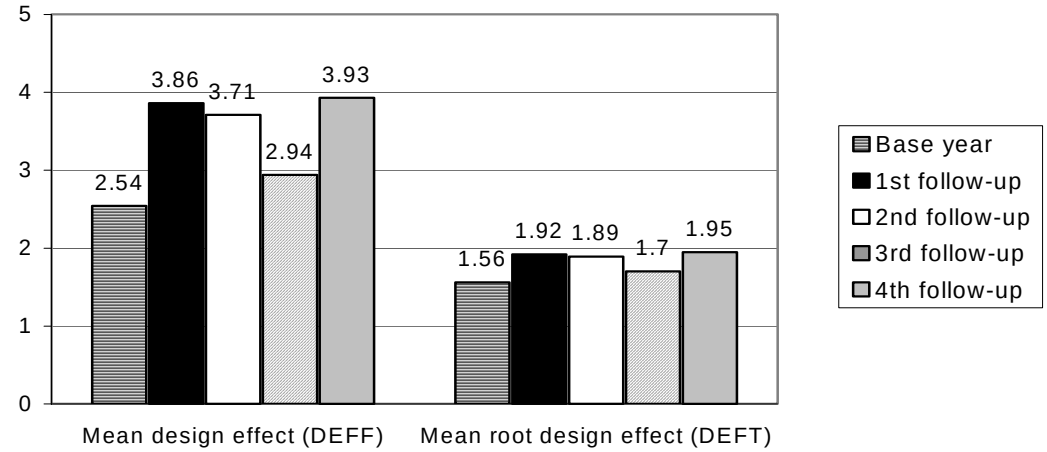
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table 3.9.7.—Mean design effects (DEFFs) and root design effects (DEFTs) for the full respondent sample (students and dropouts) for the NELS:88 base-year through fourth follow-up studies: 1988-2000

Statistic	Mean DEFFs					Mean DEFTs				
	BY	F1	F2	F3	F4	BY	F1	F2	F3	F4
Mean	2.54	3.86	3.71	2.94	3.93	1.56	1.92	1.89	1.70	1.95
Standard Deviation	1.11	1.68	1.68	0.78	1.49	0.33	0.41	0.37	0.22	0.35
Minimum	1.35	3.01	2.10	1.49	2.24	1.16	1.42	1.45	1.22	1.50
Maximum	5.01	8.46	11.12	5.77	9.31	2.24	2.91	3.33	2.27	3.05

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-2000.

Figure 3.9.7.–NELS:88 base-year through fourth follow-up mean design effects (DEFFs) and mean root design effects (DEFTs): 1988-2000



Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-2000.

As expected, the design effects for subsequent follow-up studies were somewhat higher than those of the base year (see table 3.9.4-A). This is a result of the subsampling procedures used in these subsequent follow-ups. For example, as mentioned earlier, table 3.9.4-B shows that subgroups have larger design effects compared to those in the base year. For 11 of the 12 subgroups, the first follow-up survey average design effects are larger than those for the base-year survey, regardless of whether the full or panel samples are considered.

The general tendency in longitudinal studies is for design effects to lessen over time, as dispersion reduces the original clustering. However, subsampling has the opposite effect. Indeed, subsampling increases design effects because it introduces additional variability into the weights with an attendant loss in sample efficiency, as may be illustrated by the case of the sophomore cohort of HS&B. For example, considerable subsampling of nonrespondents was done in the HS&B first follow-up, which had a rather higher design effect, 3.59, than the HS&B base year. Comparatively, more subsampling was done in the NELS:88 first follow-up, which had an overall design effect similar to, though somewhat higher than, the HS&B first follow-up (3.8 or 3.9 for NELS:88, 3.6 for HS&B).

The larger design effects (compared to NELS:88 and HS&B base years) in the NELS:88 first follow-up survey were probably due to disproportionality in strata representation introduced by subsampling. This is illustrated in the higher design effects for dropouts than for students (full sample: students, 3.86, dropouts, 4.71; panel sample: students, 4.71, dropouts, 4.70); dropouts were retained at a much higher rate (i.e., certainty, in other words *all* were retained) than students, who were subsampled at rates corresponding to their clustering in first follow-up schools.

To make a more exact assessment of the expected increase in design effects for the first follow-up sample, an additional analysis of the student data was conducted using NELS:88 base-year data. Standard errors and design effects were calculated on the base-year student respondents, using the same variables that were used in the base-year analysis, but using the first follow-up panel weight. Any magnitude of the increase in design effects in the first follow-up can be assessed by comparing the average design effect

obtained from this analysis with the design effect obtained using the entire base-year sample and the base-year questionnaire weight, *BYQWT*. This analysis yielded a design effect of 3.90 (root design effect=1.96) and supports the contention that the increase in first follow-up design effects is due to weighting necessary to accommodate the subsampling.

3.10 Developing Approximate Standard Errors from Design Effects

Researchers who do not have access to software for computing accurate estimates of standard errors (e.g., SUDAAN or Wesvar) can use mean design effects to approximate the standard errors of statistics that are based on the NELS:88 data.[56] Design-corrected 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 mean root design effect (DEFT). Or in other words, the standard errors for proportions can be estimated by:

$$SE = DEFT \times \sqrt{p(1-p)/n},$$

where p is the weighted proportion of respondents giving a particular response, n is the size of the sample, and DEFT is the mean root design effect.

Similarly, the standard error of a mean can be estimated from the weighted variance of the individual scores and the appropriate mean DEFT. The formula for this estimation is:

$$SE = DEFT \times \sqrt{Var/n},$$

where Var is the sample variance, n is the size of the sample, and DEFT is the mean root design effect.

The design effects tables for the various data collection waves of NELS:88 make it clear that the design effects and root design effects vary considerably by subgroup. For example, students from urban schools differ markedly from those attending suburban schools school during the fourth follow study; thus, it is important that the mean DEFT associated with the most relevant subgroup be used to calculate approximate standard errors for subgroup statistics.

Given the richness of the NELS:88 data, it is likely that standard error estimates may be needed for subgroups that have not been previously tabulated and reported in methodology reports and data file user's manuals. One rule of thumb may be useful in such situations: 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 typically be less affected by clustering than larger subgroups.) For example, estimates for Hispanic males will generally have smaller design effects than the corresponding estimates for the larger subgroups of 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 applies only when the variable used to subdivide a subgroup crosscuts schools. Sex

[56] For example, Tables 3.3.1-2 in Ingels et al. 1994 present the mean design effects for the NELS:88 base year. Tables 5-6 and 12-14 in that same section include the estimates, respectively, for the first and second follow-up data. Section 5.3 in Haggerty et al. 1996 describes the third follow-up design effects. Tables of mean design effects for the fourth follow-up data are included in appendix D of this report.

is one such variable, since 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 basis for the results presented here. A second rule of thumb 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. Thus, using the following formula,

$$Var(b-a) < Var(b) + Var(a)\,,$$

in which 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, analysts can use Var(a) + Var(b) as a conservative proxy for Var (b–a).

A final rule of thumb 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. Thus, this argues that using the mean root design effects noted above will provide conservative 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 simple random sample standard error is multiplied by the appropriate mean root design effect.

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 new rescaled, design effect-adjusted weight, which is the product of the inverse of the design effect and the rescaled case weight. For example, for the second follow-up full sample data, the following formula:

$$NEWWGT = \left(1 \big/ DEFF\right) \times \left[F2QWT_i \Big/ \left(\sum F2QWT_i / N\right)\right],$$

will develop an approximate new weight to deflate the obtained sample size to take into account the inefficiencies due to a sample design that is departure from a simple random sample. Using this procedure, statistics calculated by statistical programs 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 complete accounting of the sample design, this procedure is a decidedly better approach than conducting analyses that assume the NELS:88 data were collected from a simple random sample. Analysts applying this correction procedure should carefully examine their statistical software to and assess whether the program treats weights in such a way as to produce the effect described above.

3.11 Sources of Additional Information

For each round and component, the various user's manuals and methodology reports listed at the end of chapter 1 should be consulted, as well as the *NELS:88 Base-Year Sample Design Report*.

Guidance in using the NELS:88 case weights in analysis can be found on the ECB guide that is available on the NELS:88 1988-2000 ECB and on the NELS:88 Web Site at NCES.

On issues of *sample exclusion and ineligibility*, especially at the student level for the three in-school rounds, please see:

Ingels, S.J. (1996). *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years* (NCES 96-723). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Dowd, K.L., Baldridge, J.D., Stipe, J.L., Bartot, V.H., and Frankel, M.R. (1994). *Second Follow-Up: Student Component Data File User's Manual* (NCES 94-374). Washington, DC: National Center for Education Statistics.

Haggerty, C., Dugoni, B., Reed, L., Cederlund, A., and Taylor, J. (1996). *National Education Longitudinal Study (NELS:88/94) Methodology Report* (NCES 96-174). Washington, DC: National Center for Education Statistics.

On the sample design and weighting of the NELS:88 *High School Effectiveness Study*, the following sources should be consulted:

Scott, L.A., Ingels, S.J., Pulliam, P., Sehra, S., Taylor, J.R., and Jergovic, D. (1996). *NELS:88 High School Effectiveness Study: Data File User's Manual*. Chicago: NORC contractor report to NCES.

Spencer, B.D., and W. Foran. (1991). "Sampling Probabilities for Aggregations, With Application to NELS:88 and Other Educational Longitudinal Surveys." *Journal of Educational Statistics*, 16(1).

Qian, J. (1996). A Comparison of Weights Derived from Different Models—NELS:88 High School Effectiveness Study. *1996 Proceedings of the Survey Research Methods Section, American Statistical Association*.

Other sources cited in this chapter include:

Chong, E.K.P., and Zak, S.H. (1996). *An Introduction to Optimization*. New York: Wiley.

Chromy, J. R. (1981). Variance estimators for a sequential sample selection procedure. *Current topics in survey sampling*. Academic Press.

Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a Sample Frequency when the expected marginal totals are known. *The Annals of Mathematical Statistics, 11*: 427-444.

Deville, J.C., and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87: 376-382.

Folsom, R. E., Jr., and Singh, A.C. (2001). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *Calibration Estimation and Calibration-Adjusted Variance Estimation*. Research Triangle Park, NC: Research Triangle Institute.

Frankel, M.R. (1971). *Inference from Survey Samples: An Empirical Investigation.* Ann Arbor: Institute for Social Research.

Frankel, M.R., Kohnke, L., Buaonnanno, D., and Tourangeau, R. (1981) *High School and Beyond Sample Design Report*. Chicago: National Opinion Research Center.

Frase, M. (1989). *Dropout Rates in the United States: 1988* (NCES 89-609). Washington, DC: National Center for Education Statistics.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kish, L., and Frankel, M.R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36, 2-37.

McMillen, M.M., Kaufman, P., Hausken, E., Bradby, D. (1993). *Dropout Rates in the United States: 1992* (NCES 93-464). Washington, DC: National Center for Education Statistics.

Pavel, D.M., and Curtin, T.R. (1997). *Characteristics of American Indian and Alaska Native Education: Results from the 1990-91 and 1993-94 Schools and Staffing Surveys* (NCES 97-451). Washington, DC: National Center for Education Statistics.

Pavel, D.M., Curtin, T.R., Thorne, J.M., Christenson, B., and Rudes, B. (1995). *Characteristics of American Indian and Alaska Native Education.* (NCES 95-735). Washington, DC: National Center for Education Statistics.

Shah, B.V., Barnwell, B.G., and Bieler, G.S. (1997). *SUDAAN User's Manual*. Research Triangle Park, NC: Research Triangle Institute.

Zieschang, K. D. (1990). Sample weighting methods and estimation of totals in the consumer expenditure survey. *Journal of the American Statistical Association*, 85: 986-1001.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K., and Tourangeau, R. (1990). *NELS:88 Base-Year Sample Design Report* (NCES 90-463). Washington, DC: National Center for Education Statistics.

Chapter IV
Data Collection Methodology and Results

This chapter provides an overview of both the pre-data collection activities and data collection results for the five rounds of NELS:88. Because of the similarly of their data collection methodologies, our discussion of procedures considers jointly the in-school rounds of the study—the base-year, first and second follow-up studies in 1988, 1990, and 1992. Data collection with these studies included surveys of students and school dropouts, as well as their parents, teachers, and school administrators. We then consider the NELS:88 third and fourth follow-up studies that were conducted primarily by computer-assisted telephone interview (CATI), with field follow-up with nonresponding sample members. Finally, we discuss the procedures employed by the research team to collect postsecondary education transcripts for members of the sample cohort.

4.1 School Recruitment and Pre-Data Collection Activities

4.1.1 Base Year

Before the base-year data collection effort could begin, members of the research team secured commitments from the administrators of sampled schools. Several levels of cooperation were sought before school administrators were approached. The first level involved contacting key educational organizations. For example, the Education Information Advisory Council (EIAC) of the Council for Chief State School Officers was asked to give its approval for the project. Contact was also made with the National Catholic Educational Association (NCEA) and the National Association of Independent Schools (NAIS) in order to inform them of the study and to solicit their endorsements.

For public schools, the next step involved contacting the chief state school officer in each state, usually the state superintendent of education, to explain the objectives of the study and the data collection procedures, especially those for protecting individual and institutional confidentiality. Once state-level approval was obtained, contact was made with district superintendents, where a similar information exchange and approval approach was followed. Then, upon receipt of district approval, contact was made with the school principals. The research team used a similar approach, where appropriate, for private schools organized into an administrative hierarchy.[57] District and school contacts were implemented early in the fall term of the 1987-88 school year. However, for some schools, cooperation was not achieved until nearly the end of that term as the result of lengthy district-level review processes or the delayed selection of schools chosen to replace refusal schools.

Within each cooperating school, principals were asked to designate a school coordinator who would serve as a liaison between data collection staff and selected respondents, the school administrator, students, teachers, and parents. The school coordinator, who was often a guidance counselor or senior teacher, but sometimes the principal or assistant principal, handled all requests for data and materials, as

[57] For example, in Catholic school dioceses, a "courtesy" call to request permission to contact the principal of the school was completed at the administrative level before the school principal or other administrator was approached.

well as all logistical arrangements for data collection on the school premises. Included among these responsibilities was annotating the list of eligible students to identify students whose physical or mental disabilities or whose limited English proficiency would preclude their study participation. Coordinators were also asked to classify all eligible students as Hispanic, Asian-Pacific Islander, or "other" (neither Hispanic nor Asian-Pacific Islander), and to distribute parental permission forms to sampled students. (Procedures for selecting the base-year student sample are described in chapter 3.)

4.1.2 First Follow-up

Although the data collection procedures employed in the first follow-up study were closely modeled after the base-year design, the study necessitated new pre-data collection activities, in particular, student tracing. In its transition to high school, the 1988 8th-grade cohort had dispersed from 1,052 base-year participating schools to nearly 4,000 high schools by the time of tracing. Moreover, student mobility continued to each school's survey session in the spring term of the 1989-90 school year. This high degree of mobility necessitated a subsampling effort to select the first follow-up sample.

From January to June of 1989, the research team traced base-year sample members to their 1989-90 school of attendance. A student sample was selected (procedures are described in chapter 3), and state, district, and school permission to conduct the study was secured. While a few students remained in the base-year school, and more in the base-year district, most schools were new to NELS:88, as were a number of districts. From September to December 1989, all first follow-up schools were contacted again in the fall of 1989, primarily to re-verify student enrollment (both to identify transfers and dropouts), freshen the student samples, and schedule in-school data collection sessions.

4.1.3 Second Follow-up

The second follow-up survey was executed in three phases which spanned two years. Pre-data collection activities took place during phases 1 and 2, while data collection took place during phase 3.

Conducted from January to June of 1991, phase 1 of the second follow-up survey encompassed the pre-data collection activities of tracing sample members to their school of attendance and securing state, district, and school permission to conduct the study. State cooperation with NELS:88 was secured for all fifty states and the District of Columbia. District and school-level cooperation was secured for first follow-up schools with four or more sample members still in attendance in the spring of 1991.

Tracing sample members served two purposes. First, the tracing located sample members for data collection purposes, and secondly, it helped to define the schools that would be included in the second follow-up contextual component sample. In the first follow-up, 21,474 of over 26,000 base-year sample members were retained. In the second follow-up, the entire first follow-up sample was retained; however, for cost reasons, contextual data was collected only for a subset of students at approximately 1,500 schools. In tracing the first follow-up sample, the research team found that sample members had dispersed from the approximately 1,500 high schools during the 1990 school year to 2,258 schools in 1992.

To maximize the number of students for whom the full complement of contextual data—school administrator and teacher reports—and high school transcripts would be collected, the number of sampled students at each school was determined during tracing. The school sample was then drawn so that the greatest number of students would be included in the school sample. Students from schools not selected for the school sample were surveyed, but teacher and school administrator data were not collected and high school transcripts were not requested.

From September to December 1991, phase 2 pre-data collection activities occurred for all components of the study, and some phase 1 activities continued. District and school-level cooperation were gained for any schools selected for the second follow-up sample for which cooperation was not gained in phase 1. Tracing continued for sample members who were not located during phase 1, and enrollment was verified again for students who were traced to a school which was selected for the second follow-up school sample. Students attending a school not included in the second follow-up school sample and sample members who had left school were traced to their school of attendance or to a home address

4.2 Data Collection Results and Procedures, In-School Rounds

This section summarizes data collection results and procedures for the in-school waves of NELS:88 in 1988, 1990, and 1992, including procedures for collecting data from out-of-school respondents such as dropouts. Overall response rates for the three in-schools rounds are summarized below in table 4.2-A through table 4.2-C; more detailed response rate tables may be found in appendix C. In addition, table 4.2-D depicts the cross-round pattern of participation for student questionnaire and test completion, based on the base-year to second follow-up 8th-grade panel and the first follow-up to second follow-up 10th-grade panel.

Table 4.2-A.—Summary of completion rates for the NELS:88 base-year study

Instrument	Completed	Weighted (in percent)	Unweighted (in percent)
Student questionnaires	24,599	93.41	93.05
Student tests	23,701	96.53[1]	96.35[1]
Parent questionnaires	22,651	93.70	92.08
Teacher ratings of students	23,188	95.91[2]	94.26[2]
School administrator questionnaire	1,035	98.92	98.38

[1] Percentages of cases for which a student questionnaire was obtained for which a cognitive test was also obtained.

[2] Percentage of student respondents for whom at least one teacher rating was completed.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

Table 4.2-B.—Summary of completion rates[1] for the NELS:88 first follow-up study

Instrument	Completed	Weighted (in percent)	Unweighted (in percent)
Student questionnaires	18,221	91.09	94.10
Student tests	17,352	94.14[2]	95.23[2]
Dropout questionnaires	1,043	90.97	89.84
Dropout tests	522	48.56[2]	50.05[2]
School questionnaires	1,291	†	97.07
School questionnaires[3]	17,663	91.97	96.94
Teacher questionnaire[4]	15,908	80.51	87.31

† Because of student dispersal between the BY and F1 studies and required high school subsampling, school weights were not calculated.

[1] Table is based on the original (1992-93) release of the first follow-up file. The NELS:88 second follow-up re-release of first follow-up student data contains a slightly different sample number due primarily to base-year-ineligibles who were subsequently returned to the sample. Additional details about the sample sizes of these releases are in the *NELS:88 Second Follow-Up: Student Component Data File User's Manual*.

[2] Percentages of cases for which a student/dropout questionnaire was obtained for which a cognitive test was also obtained.

[3] Coverage rate for student participants of the total sample who also have a completed school administrator questionnaire.

[4] Percentage of student respondents for whom at least one teacher rating was completed.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1990.

Table 4.2-C.—Summary of instrument completion rates for the NELS:88 second follow-up study: 1992

Instrument	Completed	Weighted (in percent)	Unweighted (in percent)
Student questionnaires	16,842	91.0	92.5
Student cognitive tests	13,267	76.6[1]	78.8[1]
Dropout questionnaires	2,378	88.0	87.6
Dropout tests	959	41.7[1]	40.3[1]
School questionnaire[2]	1,326	†	97.1
School questionnaire[3]	15,409	98.3	98.2
Parent questionnaire[4]	16,395	90.6	93.2
Teacher questionnaire[5]	9,853	90.8	90.7

† Because of student dispersal between the BY and F1 studies and required high school subsampling, school weights were not calculated.

[1] Percentages of cases for which a student/dropout questionnaire was obtained for which a cognitive test was also obtained.

[2] Twelfth-grade school completion rate for school questionnaires of eligible contextual schools where at least one student has completed a questionnaire.

[3] Coverage rate for student participants of the total sample who also have a completed school administrator questionnaire.

[4] Parent completion rate is based only on those sample members who completed a student/dropout questionnaire.

[5] Percentage of student respondents for whom a teacher rating was completed.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1992.

Table 4.2-D.—Pattern of participation across the in-school rounds of NELS:88 for the 8th-grade and 10th-grade respondent cohorts: 1988-1992

NELS:88 8th-Grade Cohort									
Questionnaire completers					**Cognitive test completers**				
1988	**1990**	**1992**	**N**	**%**	**1988**	**1990**	**1992**	**N**	**%**
N	N	N	185	1.0	N	N	N	438	2.2
N	N	Y	122	0.6	N	N	Y	122	0.6
N	Y	N	146	0.7	N	Y	N	463	2.4
N	Y	Y	799	4.1	N	Y	Y	906	4.6
Y	N	N	331	1.7	Y	N	N	1,270	6.5
Y	N	Y	638	3.2	Y	N	Y	683	3.5
Y	Y	N	935	4.8	Y	Y	N	3,861	19.6
Y	Y	Y	16,489	83.9	Y	Y	Y	11,902	60.6
Total			19,645	100.0	**Total**			19,645	100.0
NELS:88 Sophomore Cohort									
Questionnaire completers					**Cognitive test completers**				
1988	**1990**	**1992**	**N**	**%**	**1988**	**1990**	**1992**	**N**	**%**
†	N	N	129	0.7	†	N	N	867	4.8
†	N	Y	293	1.6	†	N	Y	566	3.1
†	Y	N	1,005	5.5	†	Y	N	4,169	22.9
†	Y	Y	16,749	92.2	†	Y	Y	12,574	69.2
Total			18,176	100.0	**Total**			18,176	100.0

† Not applicable this wave.

NOTE: N = did not complete instrument; Y = completed the instrument.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

4.2.1 Student Survey Procedures

The procedures by which the above results were achieved are explained below. For in-school administrations, essentially the same student survey data collection methods were used for the three in-school rounds. Student questionnaires and tests were administered in group sessions. Survey administration, normally conducted in a school classroom or library, consisted of several steps. Students first completed the student questionnaire. A ten-minute break followed, during which field staff reviewed the questionnaires for missing or invalid responses to critical items.[58]

Following the break, on-site data collection personnel administered an 85-minute battery of cognitive tests. The tests consisted of four timed sections devoted to mathematics, reading, science, and

[58] A specially designated oval indicating "no retrieval" was marked whenever the missing data could not be retrieved when the respondent refused or was unable to clarify the response.

social studies (history/government). For the base-year data collection, all students completed the same test form in all four subjects. In the first and second follow-ups, a special feature of test administration was that there were multiple forms of varying difficulty. A specific form was pre-assigned to each student based on the ability estimate (theta) taken from the prior round's test score. Once the test battery was completed, an attempt was made to retrieve missing (or inappropriately marked) questionnaire items before the student left the room.

At the end of the session, arrangements were made to conduct make-up sessions for students who were unable to attend the survey day. If fewer than five students were scheduled for a make-up day, the school coordinator was asked to handle the arrangements and oversee its administration.[59] When five or more students were scheduled, or in instances where the school coordinator was unavailable to conduct a make up, field data collection staff returned to the school.

While the same in-school student data collection methodologies were used for the NELS:88 base-year and first and second follow-up studies, the first and second follow-up studies also collected data from cohort members in out-of-school sessions. Students who were not enrolled in sampled schools, who had missed in-school data collection sessions, or who were enrolled in schools that had refused to participate in the study were invited to off-campus sessions and administered the student questionnaire and cognitive tests. Dropouts were also asked to attend these sessions and were surveyed alongside sample members who were currently enrolled in school. One to three sample members typically attended off-campus survey sessions. If a sample member was unable to attend an off-campus group survey session, the individual was surveyed either over the telephone or in person. Cognitive test data were not collected during these telephone interviews.

4.2.2 First and Second Follow-up Dropout Surveys

After the base year of NELS:88, the research team carefully monitored the enrollment status of all sample members. Individuals who dropped out of school before data collection received a dropout questionnaire rather than a student questionnaire.

Definition of a dropout. For the purposes of the first follow-up data collection, the following definitions were used to identify sample members who dropped out of school:

1. An individual who, during the spring of 1990, according to the school (if the sample member could not be located), or according to the school and home, was not attending school or, more precisely, had not been in school for four consecutive weeks or more and was not absent due to accident or illness, or

2. A student who, during the spring of 1990, had been in school less than two weeks after a period in which he or she had missed school for four or more consecutive weeks not due to accident or illness.

Because contact was made with the schools during each of the four phases during the first follow-up, the enrollment status of each sample member was collected at four separate time periods. If at any point in phases 1–4 of data collection, a sample member met the above criteria, he or she was considered a dropout.

[59]To ensure respondent confidentiality, school coordinators were prohibited from reviewing student questionnaires for completeness. Instead, data collection supervisors carried out the review and retrieved problem data by telephone.

Some sample members who were initially identified as dropouts later re-enrolled in their school before data collection took place in phase 3. A student in this situation was no longer considered a dropout, but instead was classified as a "stopout." Stopouts were defined as a student who had a dropout episode between spring term 1988 and spring term 1990, but who were back in school in the spring term of 1990. At the data collection level, stopouts who were identified in phase 1 or phase 2 as a dropout, but who, in phase 3, had been attending school for two weeks or more were administered the first follow-up student questionnaire and cognitive test battery. Stopouts who had been attending school for less than 2 weeks were administered the dropout questionnaire.

When a school official identified a sample member as a dropout, interviewers were instructed to contact the household to confirm the status of the sample member. If either the sample member or an adult household member indicated that the dropout definition above was applicable, the sample member was classified as a dropout. This policy of confirming status through the household was applied during all four points of enrollment status verification.[60]

Furthermore, whenever a sample member was identified as a dropout, the sample member was flagged as such, and the date he or she dropped out of school recorded. If subsequent enrollment verification contacts revealed that the sample member had returned to school, the date he or she returned was recorded. Once a sample member was flagged as a dropout, regardless of whether or not he or she returned to school, the flag was maintained.

The NELS:88 second follow-up dropout survey sought to interview all sample members who left school prior to graduation, including both first follow-up dropouts who had not returned to school and sample members who dropped out after the first follow-up. All sample members appear on the second follow-up student data file regardless of their spring 1992 enrollment status. Basic classification variables and test data appear for both students and dropouts, though dropout questionnaire data appear separately on the ECBs for these data collection waves.

School enrollment classification and data collection. In order to determine which sample members were eligible to complete a dropout questionnaire, school enrollment status was determined for all sample members during the spring of 1992.

Four enrollment categories were identified. The first category included high school students who were enrolled in a school culminating in a high school diploma. These students were administered the student questionnaire and, when possible, the cognitive test battery. Early graduates were included in this category, and were asked to report retrospectively on the school from which they graduated and to complete supplemental questions about their reasons for graduating early.

The second category encompassed sample members who dropped out of high school but later re-enrolled in a high school program to obtain a high school diploma. These sample members were administered the student questionnaire and, when possible, the cognitive test battery.

The third category contained sample members who dropped out of high school but subsequently pursued an equivalent to a high school diploma, usually the General Educational Development test (GED). If an alternative completer had finished the requirements of his or her equivalency program (e.g. passed GED test), the individual was classified as a "completer" (in effect, an early graduate by alternative means) and the student questionnaire (including the early graduate supplement) was

[60]When schools identified a sample member as a dropout but the sample member or a household member identified the person as a student, information about the student's new school of enrollment was collected and the school contacted to verify the student's enrollment status.

administered. If the alternative completer had not yet fulfilled the requirements for certification, the sample member was administered a dropout questionnaire. In both cases, the cognitive test battery was also administered when possible.

Dropouts constituted the fourth enrollment category. These sample members had left their high school by the spring of 1992 and were not working toward an alternative certification. Dropouts were administered a dropout questionnaire and, when possible, the cognitive test battery.

Regardless of whether a dropout completed a student or dropout questionnaire, data collection efforts for the dropout component of the second follow-up were similar to those in the first follow-up survey. Interviewers attempted to survey most dropouts in off-campus survey sessions with testing conditions similar to in-school sessions.

4.2.3 School Administrator Survey

In all three in-school rounds, the school principal or headmaster was asked to complete a self-administered questionnaire. Questionnaires for school administrators who did not initially return their completed questionnaire were collected through telephone follow-up. As noted above, however, in the second follow-up study, the school sample for contextual data was as a subset of the NELS:88 schools and only those administrators were contacted.

4.2.4 Teacher Surveys

A self-administered teacher questionnaire was distributed to selected teachers of the sampled students. In the base year, each school was randomly assigned to one of the following combinations of curriculum areas: mathematics and English; mathematics and history; science and English; and science and history. In each NELS:88 school, data were collected from each sampled student's current teacher(s) in the two designated subject areas. This selection procedure was designed to ensure representation of mathematics or science curriculum and English or history in all schools. Combinations of English and history as well as science and mathematics were excluded by the design. The design also achieved balanced representation of the four curriculum area combinations across the school variables of control (public, Catholic, and other private); level (elementary, middle, junior-senior high school); geographical area; and school size.

In the NELS:88 first follow-up teacher survey, up to two teachers of each first follow-up sample member were asked to complete a self-administered teacher questionnaire. To maximize the longitudinal comparability of teacher data, NELS:88 first follow-up teachers for each student were selected in the same subject combinations as in the base year: mathematics-English, mathematics-history, science-English, or science-history. Freshened students who were not enrolled in the 8th grade in the base year, and hence had not been assigned a subject combination previously, were assigned the subject combination of their base-year "linked" partner.

The teacher survey was repeated in the NELS:88 First Follow-up (1990). In some situations a teacher report was collected in a subject area other than the student's assigned subject combination. If a student was not enrolled in classes in his or her assigned subject area, then a teacher report was collected in another one of the four subject areas. If a student was enrolled only in one of the four subject areas, then only one teacher report was collected for the student. Additionally, the subject area of the student's teacher report was sometimes substituted with another subject area in order to reduce the burden of the teacher survey on teachers who were asked to report on eight or more NELS:88 students. Possible student-teacher subject pairings in the base year and first follow-up are presented below. Same-subject pairings (e.g., English–English) pertain to situations in which different teachers instructed the sample

member in the same subject but different courses, or where the same teacher instructed the sample member in two different courses of the same subject matter.

Base Year	**First Follow-up**
English. ..... Mathematics	English..... Mathematics
History…… Mathematics	History ..... Mathematics
Science... ..... History	Science ..... History
Science. ..... English	Science ..... English
	Science ..... Mathematics
	English..... History
	English..... English
	History ..... History
	Mathematics ..... Mathematics
	Science ..... Science

In the second follow-up teacher survey, one teacher report was collected for each student attending a NELS:88 school who was enrolled in a mathematics or science class. For students enrolled in both a mathematics and a science class, only one teacher report was collected. For these students, the subject area of the second follow-up teacher report was the same as that of the student's base-year teacher report. Some second follow-up freshened students, who had no base-year subject assignment, were also enrolled in both a mathematics and a science class. For these freshened students, the subject area of the teacher surveyed in the second follow-up was the same as the base-year subject area of the student's linked partner in the freshening procedure.

Because students were surveyed at NELS:88 schools from January 1992 through the end of the 1991-1992 academic year, self-administered questionnaires were mailed to teachers in two mailings depending on when the students at the school were surveyed. Teachers at schools at which the students were surveyed before April 1, 1992, were mailed a questionnaire in early February 1992. Teachers at schools at which the students were surveyed on or after April 1, 1992, were mailed a questionnaire in early March 1992.

The base-year and first follow-up teacher surveys had sought reports from the spring term teacher. This procedure was somewhat modified in the second follow-up. For most students, a teacher report was collected from the fall term teacher in the selected subject. However, if the students at a school was surveyed on or after April 1, 1992, then the teacher questionnaire was mailed to the selected subject's spring term teacher of the selected subject for the student. This design was based on the assumption that early in the spring term, the fall term teacher was the most familiar with and could most fully assess the student, and in many cases, the fall and spring term teachers were the same. After April 1, a teacher report was collected from the spring term teacher because at that time the spring term teacher was more likely to have had sufficient interaction with the student to make a full assessment of the student in the teacher questionnaire, and the fall term teacher might have difficulty recalling a student he or she had not instructed in several months. Interviewing the spring term teacher for students interviewed in school data collection sessions after April 1 also provided better articulation with the student cognitive tests than interviewing the fall term teacher in late spring.

4.2.5 Base-Year and Second Follow-up Parent Surveys

During the base-year study, a self-administered questionnaire was hand-delivered by each sampled student to his or her parent or guardian. The questionnaire included a written request that it be completed by the parent or guardian most familiar with the student's current school situation and educational plans.

Following telephone prompting of nonresponding parents, interviewers attempted to administer the parent questionnaire over the telephone. If an interviewer was unable to complete the interview over the telephone, the he or she made a personal visit to the respondent to conduct a face-to-face interview.

In the second follow-up, a self-administered, forty-minute questionnaire was collected from a subsample of parents or guardians of NELS:88 students. The initial parent questionnaire mailout took place in May 1992. Like the base-year parent survey, instructions in the questionnaire and accompanying letter directed the parent or guardian who was most knowledgeable about the teenager's current school situation and educational plans to complete the questionnaire. In accordance with these instructions, the respondent was self-selected.

Whereas the base-year parent survey asked parents to complete the questionnaire near the same time the student was interviewed, the second follow-up instrument included questions about postsecondary educational costs which precluded an exact temporal correspondence between the administration of the two surveys. Because financial aid decisions are frequently not received until late in the spring of the teenager's 12th-grade year, the parent questionnaires were mailed in May 1992, to ensure that the parents and guardians would be able to answer these questions fully. For parents who completed the interview after the end of the 1991- 1992 academic year, the parent questionnaire instructed parents to refer to the spring of 1992 when answering questions about the teenager's school life.

The parent instrument was designed as a self-administered questionnaire, but many parents completed the survey over the telephone with an interviewer. To minimize any differences between the two modes of administration, interviewers were trained to adapt the questions to make sense when asked over the telephone. Interviewers also encouraged parents to read along in the questionnaire if they had a copy at hand.

4.2.6 Dropout Survey

Data collection for the dropout survey was executed during two data collection periods (January to July 1990 and January to June 1991). During the initial data collection period, interviewers administered the dropout questionnaire and cognitive tests to members of the dropout cohort during off-campus group administration sessions, described in section 4.2.1.

During the second dropout data collection period, a different data collection effort took place. In an attempt to obtain a more precise estimate of the cohort dropout rate for the 8th-grade class of 1988, enrollment status information was gathered for nonrespondents, who had been previously identified as dropouts (sample members who were identified as dropouts by school officials but not home-confirmed), and base-year ineligible students.

4.2.7 Survey of Base-Year Ineligible Students

The Base-year Ineligibles (BYI) Study of the NELS:88 first follow-up was a followback of students who had been excluded because of linguistic, mental, or physical obstacles to participation when the baseline sample of 8th-graders was drawn in the 1987-88 school year. The BYI study had several purposes, the primary foci of which were to correct for potential sample undercoverage; to accommodate the group of 1988-ineligible sample members who were 1990-eligible sophomores, and hence must be added to the 1990 survey to ensure its cross-sectional representativeness; and to provide a basis for a corrected cohort dropout estimate taking account of both 1988-eligible and 1988-ineligible 8th-graders two years later.

Eligibility information for the 1990 wave was successfully gathered for 93.9 percent of the sample of excluded base-year cohort members. For excluded students who were identified as eligible, student or dropout questionnaires were administered either in-person or over the telephone. Cognitive tests were administered to a small percentage of these students. For students who remained ineligible, school enrollment status and other key characteristics were obtained.

In the second follow-up an attempt was again made to reassess the eligibility status and ascertain the enrollment status of students who: 1) had been excluded because of linguistic, mental, or physical obstacles to participation when the baseline sample of 8th-graders was drawn in the 1987-88 school year, were subsampled into the Base Year Ineligibles Study in the first follow-up, and were ineligible for the first follow-up survey; 2) were eligible in the base year but became ineligible in the first follow-up; or, 3) were identified as ineligible when selected through the freshening process in the first follow-up. Eligibility information was gathered for 94.7 percent of the excluded sample members. For excluded students who were identified as eligible, second follow-up student or dropout questionnaires were administered either in-person or over the telephone. Cognitive tests were administered to a small percentage of these students. For students who remained ineligible, school enrollment status and other key characteristics were obtained. For eligibility and completion rate data, see appendix C. For details about the ineligibles study, see *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status after Four Years* (Ingels 1996; NCES 96-723).

4.2.8 High School Effectiveness Study

Data collection for the baseline of the High School Effectiveness Study (HSES), an independent component of NELS:88, was conducted concurrently with the NELS:88 first follow-up. The HSES and NELS:88 first follow-up school samples overlapped to a high degree, as did the student samples to a lesser extent. Data collection instruments and procedures for the HSES baseline were almost identical to those used in the NELS:88 first follow-up. NELS:88 data users should note that HSES data are not included in the NELS:88 base-year through second follow-up ECB release, but rather, must be requested separately, from NCES.

In the 247 participating HSES schools, HSES sample members were administered the NELS:88 student questionnaire and cognitive test battery. If HSES students missed their scheduled in-school data collection session, they were surveyed at an off-campus survey session. Unlike the NELS:88 first follow-up, HSES sample members who were no longer attending the HSES school at which they were sampled were not pursued or surveyed; however, enrollment status for these sample members was gathered from their original HSES school. School administrator and teacher data were gathered for HSES students using NELS:88 first follow-up instruments and procedures. Data collection for the follow-up wave of the High School Effectiveness Study (HSES) was conducted concurrently with the NELS:88 second follow-up. The HSES and NELS:88 second follow-up school samples overlapped to a high degree, as did the student samples to a lesser extent. Data collection instruments and procedures for the HSES were the same as those used in the NELS:88 second follow-up.

In 246 of the 247 schools participating in the baseline (one HSES school closed between the baseline and the followback), HSES sample members were administered the NELS:88 second follow-up student questionnaire and cognitive test battery. If HSES students missed their scheduled in-school data collection session, they were surveyed at an off-campus survey session. Like the HSES baseline, HSES sample members who were no longer attending the HSES school at which they were sampled were not pursued or surveyed, but their enrollment status was collected from their original HSES school. Parent, school administrator and teacher data were gathered for HSES students using NELS:88 second follow-up instruments and procedures.

In the 1992 round of HSES transcripts were collected and processed for all sample members eligible for the baseline or followback. Course offerings documents for the 1991-92 school year were also collected from HSES schools and used in transcript coding. Unlike the NELS:88 second follow-up, school-level and course-level data were also abstracted from the course catalogs and other documents provided by HSES schools. When used with transcript data for HSES sample members, course offerings data facilitate the investigation of coursetaking patterns by student characteristics and the relationship of these patterns to student outcomes. The data also allow for more fine-grained analysis of learning opportunities because the data are informative of all the courses offered at a school during the 1991-92 academic year. A detailed discussion of the data collection procedures for the High School Effectiveness Study is provided in Scott et al., the *NELS:88 High School Effectiveness Study: Data File User's Manual*.

4.2.9 High School Transcript Component

In August 1992, transcript survey materials were mailed to the principals of the NELS:88 and non-NELS:88 schools attended or most recently attended by sample members eligible for the survey. (The sample for the transcript component comprised all eligible NELS:88 second follow-up sample members who were: 1) students enrolled in NELS:88 schools; 2) early graduates, regardless of school affiliation; or 3) dropouts [including GED recipients]. Sample members who were ineligible for the base year, first follow-up and second follow-up and were enrolled in the 12th-grade in 1992 were also part of the sample.) Because of the variability in transcript format across schools, explicit instructions for transcript preparation were provided. School staff were asked to retrieve from alternate sources any data elements that were not included on the school's transcripts. Transcript preparers were also asked to note any in-school survey session day transfers on survey documents, to facilitate the pursuit of additional records from transfer schools.

Two weeks after survey materials were mailed, nonresponding principals were prompted for the return of transcripts with a postcard reminder. Principals who did not return transcripts within 3 weeks of the postcard prompt were prompted over the telephone. Telephone prompting of nonresponding principals continued from October 1992 to February 1993. Field visits to schools requesting assistance in the preparation of transcripts were conducted in February and March.

Abstraction of student- and course-level data from transcripts began in October 1992 and continued through March 1993. Retrieval of missing critical items from school staff occurred concurrently. Coding of transcript courses began in November 1992, and continued through April 1993. Courses were coded using the course catalog for the school or district, in accordance with the Classification System of Secondary Courses, updated for the 1990 NAEP High School Transcripts Study. When a school or district catalog was unavailable, courses were coded by title alone. Further information about data collection for the high school transcript component is provided in Ingels et al., *NELS:88 Second Follow-Up: Transcript Component Data File User's Manual* (NCES 95-377).

4.3 Data Collection Results and Procedures for the Out-of-School Rounds

This section summarizes the data collection results and procedures for the two out-of-school rounds for NELS:88 conducted during 1994 and 2000. Overall response rates for the two rounds are summarized in table 4.3; more detailed response rate tables for the NELS:88/94 and NELS:88/2000 follow-up studies are provided in appendix C, including important sample subgroups such as race-ethnicity, sex, socioeconomic status, dropout status, and type of original school.

Table 4.3.—Summary of completion rates for the NELS:88 third and fourth follow-up studies: 1994-2000

Instrument	Selected	Completed	Weighted (in percent)	Unweighted (in percent)
Third Follow-up Study	15,875[1]	14,915	90.9	94.0
Fourth Follow-up Study	15,649[2]	12,144	82.7	77.6

[1]This does not include 89 ineligible or deceased sample members.

[2] The total sample size excludes 315 deceased, incapacitated, or otherwise out-of-scope cases.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1994-2000.

After 1992, most sample members had left high school and made the transition to the workforce or postsecondary education. This dispersed out-of-high-school sample was best studied through a different data collection methodology: computer-assisted telephone interview (CATI), supplemented with self-administered and field interviewer-administered surveys or interviews.[61] The two predominantly CATI rounds are described in the sections below, along with the postsecondary education transcript study that followed the fourth follow-up interview in late 2000.

The third and fourth follow-up surveys collected a fourth and fifth wave of data from the 8th-grade cohort of 1988, approximately two and eight years, respectively, after the majority completed high school. Interviewing took place from February through June 1994 for the third follow-up study, and from January through September 2000 for NELS:88/2000. Notably, both studies conducted some data collection activities (e.g., advance locating and mail contact with sample members and other contacts) prior to the start of interviewing. The principal mode of data collection in both follow-ups was CATI, supplemented with a self- or field-administered hard-copy questionnaires in the third follow-up study and computer-assisted personal interviews (CAPI) in the fourth follow-up for sample members who were unwilling or unable to complete the CATI interview or who could not be located without on-site field tracing.

4.3.1 Locating

Locating activities for the third and fourth follow-ups were quite similar and involved three different but interrelated locating activities: advance locating conducted before the start of interviewing, intensive locating (typically provided during CATI as needed and just before the start of field operations), and field locating conducted on site by specially trained field personnel.

During earlier rounds of NELS:88 data collection, locating information for sample members was collected as part of the data collection activities with the sample members and their parents. The locating information included the sample members' home and school addresses and telephone numbers; the addresses and telephone numbers of parents, other relatives, and friends of sample members; drivers license and Social Security numbers for the sample members and sometimes their parents; and information about high school and postsecondary schools the students had attended. This information was entered into a secure "NELS:88 locator database," and this information served as the starting point for advance locating activities.

[61] The third follow-up field interviewers administered hard-copy questionnaires to the sample members. Field interviewers during the fourth follow-up interview used laptop computer-based CAPI technology to administer the same instrument used by CATI interviewers.

Locating for the third follow-up study began with telephone and address updates for the sample members provided by commercial databases. For the fourth follow-up study, this initial step was conducted for both parent and sample member addresses. The locator database was also further supplemented when the addresses were submitted to Telematch, a commercial database that maintains individual address and telephone number changes, to obtain updated telephone numbers. At that time, advance letters were mailed to sample members and their parents and/or other contacts to update the sample member addresses and to gain cooperation by explaining the purposes of the study.[62] For both studies, just before the start of data collection, a standard lead letter was mailed to sample members to explain the study's purpose, inform them of the upcoming interview, and obtain additional postal service address updates.

While many of the sample members were found using information collected in the prior follow-up study, certain subgroups in NELS:88/94, notably nonresponders and poor responders in previous rounds, dropouts, and American Indians/Alaska Natives, required additional resources. For the third follow-up study, a two-tiered tracing process was used to locate those sample members. The first tier, a general locating process, consisted of telephone calls to Directory Assistance and next-of-kin and other contacts nominated by the respondents in prior rounds of data collection. The second tier was a more specialized locating that utilized commercial locating databases and other locating sources. For the fourth follow-up study, the research team followed a slightly different approach; the "dead-ended" cases—where the trail of the student was lost—were reviewed by a tracing specialist and submitted to either FastData (an on-line database of names, addresses, and telephone numbers) or to RTI's specialized tracing operations unit (TOPS) for intensive locating. TOPS tracers had real time access to databases that contained current address and telephone listings for the majority of consumers with credit history. In addition to these credit history databases, the tracing specialists used other information sources, such as dataminers, commercial list-houses, and national change of address (NCOA) databases for information on the sample members. These sources searched for name, address, neighbor, business, telephone number, and status as deceased, incapacitated, or military personnel.[63] The fourth follow-up also employed external locating for cases that were not located during the intensive locating process, including batch submissions to FastData and the departments of motor vehicles (DMVs) in selected states,[64] and a search of the U.S. Department of Education's National Student Loan Data System (NSLDS), a directory of student financial aid applicants and recipients. Figure 4.3.1-A provides a schematic of the advance locating for the fourth follow-up study; figure 4.3.1-B presents information about intensive locating activities.

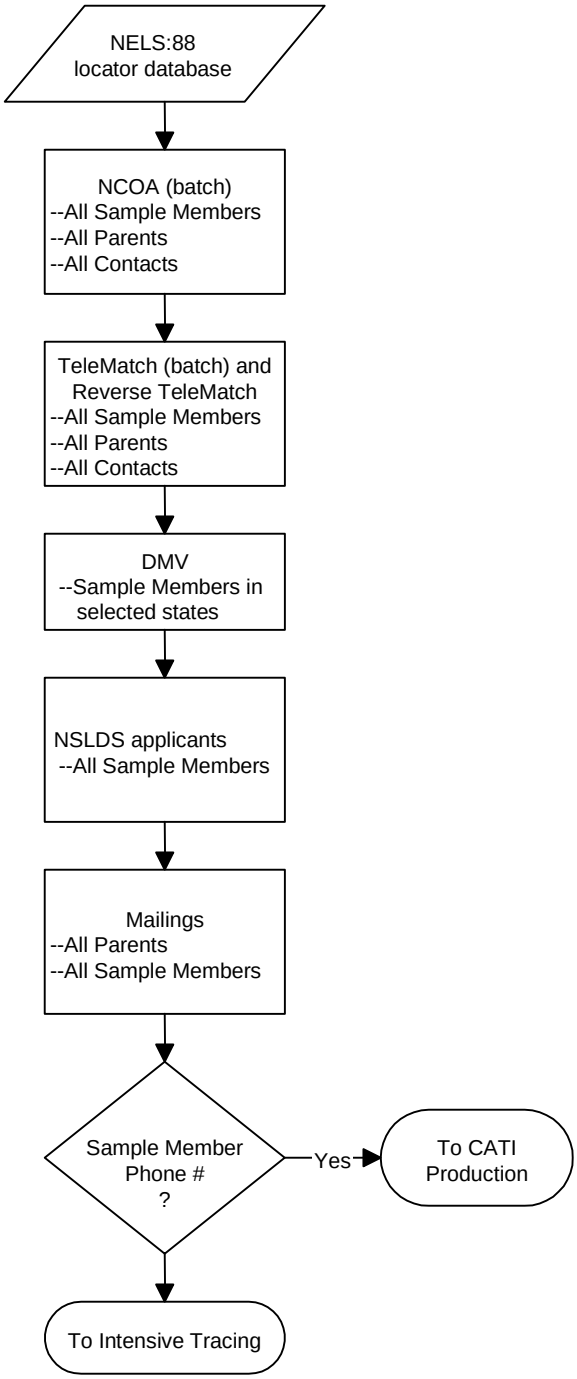
Uncompleted CATI cases were assigned to specially trained field personnel for locating and interviewing. Using local knowledge and a wide variety of tracing resources, field staff traced cases that were believed to reside in their geographic area. Primary tracing sources included: current or former neighbors, the former postsecondary and high school attended by the sample member, past or present

[62] It is important to remember that, unlike the third follow-up study, which took place two years after the previous follow-up, NELS:88/2000 occurred six years after the previous study. These were highly mobile years for much of the sample, many of whom relocated after finishing their postsecondary education, married, started and changed jobs, and the like.

[63] During pre-CATI advance locating and intensive locating during data collection, 6,753 sample members (44.3 percent of the 15,237 cohort members selected for the fourth follow-up) received some type of TOPS tracing. Some cases were traced multiple times.

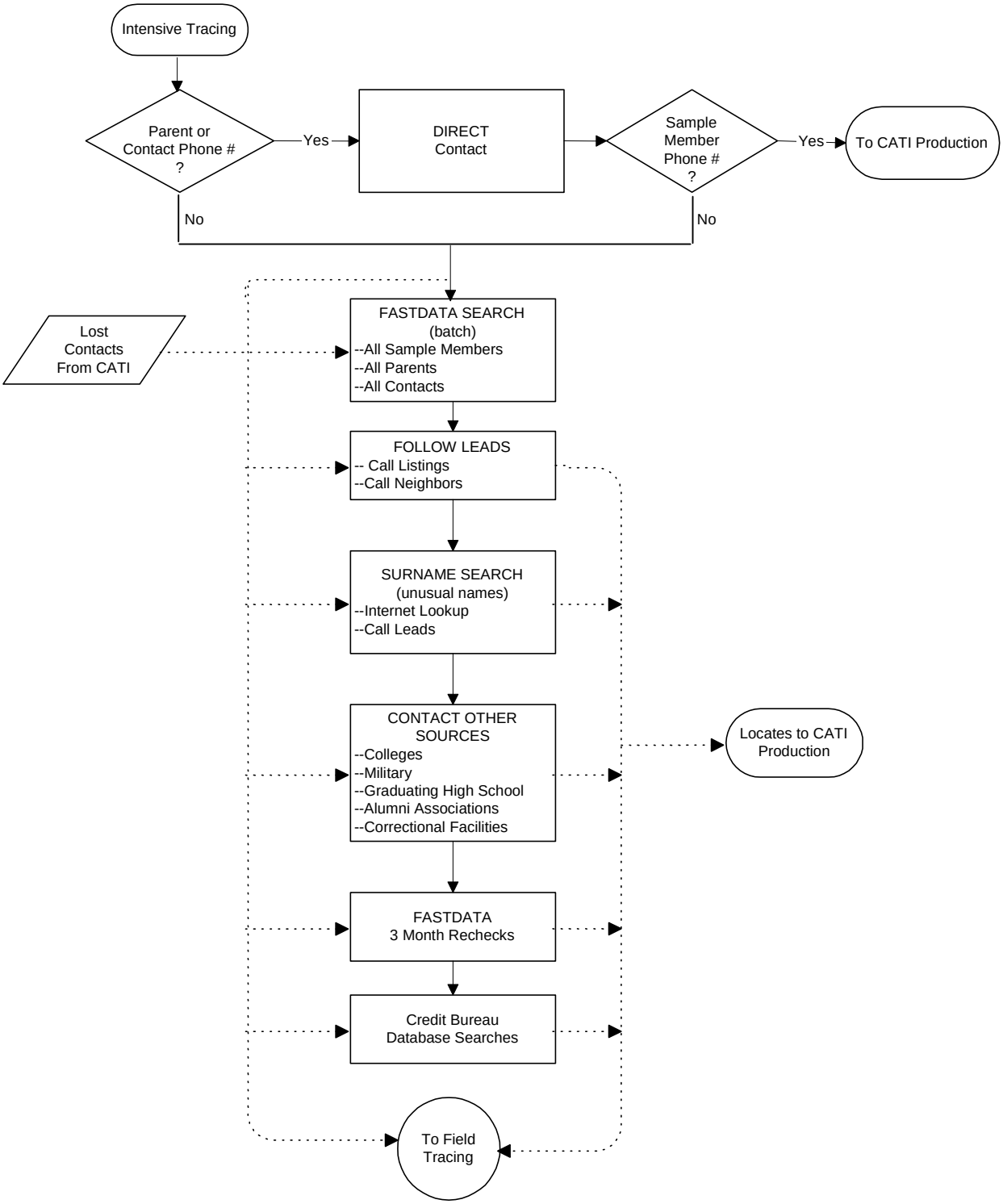
[64] Before the start of data collection, address updates on sample members were received from the DMVs in 17 states: California, Florida, Georgia, Illinois, Iowa, Mississippi, Missouri, New Jersey, New York, Ohio, Pennsylvania, Texas, Utah, Virginia, Washington, Wisconsin, and Wyoming.

Figure 4.3.1-A.—Advance locating for the NELS:88 fourth follow-up study: 2000



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Figure 4.3.1-B.—NELS:88 fourth follow-up study intensive tracing activities: 2000



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

employers, social service agency records, and government offices. Secondary tracing sources included directory assistance, public libraries, U.S. Postal Service, and state departments of motor vehicles.

For NELS:88/94, field interviewers were assigned cases not completed in CATI and, for those cases that were successfully traced, either interviewed the individuals using a hard-copy interview guide or asked them to complete a hard-copy questionnaire. The fourth follow-up study employed a slightly different model of field data collection that employed both field ***interviewers*** and field ***locators***. For this follow up, the specially trained field interviewers who located and interviewed sample members that were thought to reside in their geographic areas, were supplemented by field locators, who provided coverage in areas of the country (e.g., nonmetropolitan areas) where large clusters of students were not located.[65] These field locators, pulled from RTI's national pool of experienced field personnel, were asked simply to locate sample members and to encourage them to call the CATI center to complete telephone interviews.

4.3.2 Training of Interviewers

Interviewer training was designed to maximize the trainees' active participation and consisted of a mixture of lecture, demonstration, and hands-on practice. Interviewers received information about the study and were trained to avoid and convert refusals by reviewing the questions raised most often by respondents. For the fourth follow-up, the interviewers were also trained to perform online coding of the industries/occupations, majors/fields of study, and postsecondary educational institutions reported by the sample members.

For the third follow-up, field interviewers were trained on a flow basis. Training materials were sent to the field interviewer for self-study, with subsequent training provided by field managers. Training for the NELS:88/2000 field personnel also involved a remote in-person training session for some of the field personnel. For example, the training received by field locators (who traced but did not actually interview sample members) was comparable to the third follow-up field interviewer training. However, the fourth follow-up field interviewers (who conducted interviews with a laptop-based CATI/CAPI instrument) received intensive in-person training from project personnel at RTI's offices in North Carolina. Three-day training sessions for field interviewers were offered in March, April, and June 2000, for 71 field interviewers. A total of 7 three-day training sessions were offered for NELS:88/2000 telephone interviewers in January and February 2000.[66]

4.3.3 Telephone Interviewing

Using sample member contacting information pulled from the NELS:88 locator database, as updated during advance and intensive locating, telephone interviewers made calls to sample members. When the sample members could not be identified using this information, CATI interviewers employed additional calls to the sample member's parents and other contacts to locate the individual.

An automated call-scheduler assigned cases to interviewers based on time of day, day of week, appointment setting, and type of case considerations. Scheduler case assignment was designed to maximize the likelihood of contacting and interviewing sample members. After a certain number of call attempts (20 in NELS:88/94 and 10 in NELS:88/2000), the case was flagged for supervisory review. Depending on the circumstances, some of these cases were called by refusal converters, while others were referred for nonresponse incentives (in the fourth follow-up study), intensive locating and in-person field

[65] During the study, a total of 71 field interviewers were assigned to 40 geographic clusters; 126 field locators provided coverage in other areas of the country.

[66] This large number of sessions was required to train sufficient numbers of CATI supervisors and data quality monitors and day, night and weekend shift telephone interviewers for the study.

data collection. For the fourth follow-up interview, supervisors also authorized and arranged for calls outside of typical call center hours for interviews.[67]

Once located, some cases required special treatment. In third follow-up, interviewers entered detailed notes describing the sample member's reason for not participating and a personalized letter addressing the specific objection was sent to the sample member.

For the fourth follow-up study, a nonresponse incentive program, tested during the NELS:88/2000 field test and approved for use with the full-scale study by the federal Office of Management and Budget (OMB), was implemented to encourage sample member participation and to reduce the potential for nonresponse bias. Following the initial wave of CATI calls to sample members, three groups of "nonrespondents" received incentives. Nonresponse types included (1) sample members who refused to be interviewed for the study,[68] (2) persons who were "hard to reach" (e.g., unavailable for interviews after 10 or more telephone calls during a three-week period or who repeatedly broke CATI appointments), and (3) sample members who could not be located or contacted by telephone (e.g., their telephone numbers were unlisted or their telephone service was discontinued). Nonresponding sample members within these conditions received a $20 incentive to complete the interview. These sample members received a personalized letter delivered by overnight delivery service.[69] The letter addressed the most frequent questions or concerns raised by nonrespondents about the study. Also enclosed with the letter was a $5 bill and instructions for completing an interview by calling a toll-free telephone number. After successfully completing the interview, whether by a call-in to the toll-free number or through a subsequent call from a telephone interviewer, each respondent received an additional payment of $15 by personalized check. Cases subsequently assigned to field interviewers remained eligible for this incentive, which was delivered in cash by the field interviewer following the interview.

4.3.4 Field Interviewing

Field interviewing procedures included attempts to locate, gain cooperation from, and interview sample members. Field operations were handled differently in the third and fourth follow-up studies. NELS:88/94 used field interviewers primarily to locate sample members. Once located, the sample member either completed the interview with a telephone interviewer or completed a self-administered or field interviewer-administered hard-copy questionnaire. Similar to this approach, fourth follow-up study field locators traced sample members and encourage them to call the CATI center to complete the interview. If the interview was not completed at that time, the locator forwarded contact information for the sample member to the telephone survey unit for CATI follow-up. Field interviewers for the fourth follow-up study, on the other hand, not only located sample members but completed interviews using the same CATI/CAPI instrument used by telephone survey personnel.[70] During the course of data collection for the fourth follow-up study, 71 field interviewers served in 40 geographic clusters. These interviewers were assigned 1,676 field cases that could not be completed in CATI. A total of 126 field locators were hired to handle 749 unclustered cases.

[67] For example, foreign calls were attempted when home telephone numbers were known and the sample member communicated interest in the study (e.g., by email).

[68] Two refusals from sample members were required; multiple refusals from "other sources" such as spouses and roommates were required. Hostile refusals were not incentivized.

[69] Packages to post office boxes received overnight delivery from the U.S. Postal Service's Express Mail; all other valid addresses received letters via Federal Express.

[70] Field interviewers connected regularly to RTI's field systems group via the Internet to communicate with project personnel and to receive new assignments. Completed interviews were maintained on their laptops until that time.

4.3.5 Postsecondary Education Transcripts Study[71]

The NELS:88 Postsecondary Education Transcript Study (PETS) was carried out at the conclusion of CATI and CAPI data collection for the fourth follow-up study. Data collection for this specialized substudy began on September 5, 2000,[72] and over the next five months project staff requested transcripts from 3,213 postsecondary institutions that NELS:88/2000 respondents reported attending during ***either*** the NELS:88/94 or NELS:88/2000 studies. The study was designed to provide reliable and objective information about the types and patterns of postsecondary courses taken by NELS:88 sample members and to supplement the large NELS:88 database of factors that may predict or explain student postsecondary education and economic outcomes.

In an effort to reduce burden on the postsecondary institutions, data collection procedures for PETS were designed to follow, where possible, each institution's typical procedures for producing and distributing student transcripts. When requested by school officials, institutions were reimbursed for the transcripts and catalogs; and university officials were asked to provide transcripts in the most convenient format.[73] Records were obtained from all types of postsecondary institutions: schools offering only short-term vocational programs, as well as large land grant universities with separate graduate and professional schools. Returned transcripts and related school catalogs and bulletins were inventoried, transcript identification numbers affixed to each, and unique identifying information removed.[74] At that time, data entry and coding of the transcripts took place. Information from the transcripts, including terms of attendance, fields of study, specific courses taken, and grades and credits earned, is currently being coded and processed into a system of data files that will supplement the base-year through fourth follow-up data. This information will be released as a restricted-use ECB containing only the postsecondary transcript information.

Postsecondary institutions attended by NELS:88/2000 respondents. Only a subset of postsecondary institutions identified by NELS:88/2000 respondents were approached for student transcripts. To be eligible for the study, the "postsecondary education institutions" reported by the NELS:88 sample members during the third and fourth follow-up studies were required to be identified in NCES' Integrated Postsecondary Education Data System institutional characteristics file for the 1997-98 academic year (IPEDS-IC 1997-98). Thus, student-reported foreign schools, military training programs, and other non-credit granting institutions not included in the IPEDS-IC file were excluded from transcript data collection. IPEDS institutions that had closed, merged with other schools, or changed names and locations were retained, and every attempt was made by data collection personnel to collect transcripts from the schools or their successors.

A total of 3,213 unique postsecondary institutions from the 1997-98 IPEDS-IC file, representing some 16,020 transcripts, were reported by the fourth follow-up respondents. Table 4.3.5-A provides a description of the characteristics of the institutions reported by the NELS:88/2000 respondents. Approximately half of the schools reported by the sample members were public, 4-year institutions.

---

[71] Data collection for PETS involved two phases. Phase I is described here. Phase II involved 401 student transcripts from 256 schools that had not been reported during the NELS:88 interviews. Ninety-two percent of these transcripts were subsequently collected using procedures comparable to those described for phase I.

[72] Data collection for the nonresponse subsample with 386 members of the NELS:88 cohort was underway when transcript data collection commenced. These respondents were subsequently added to the transcript control system.

[73] While most institutions provided official, signed transcripts, many sent unofficial versions (advising forms) of the documents. Institutions could also download electronic versions of the transcripts to a secure FTP site.

[74] Student names, dates of birth, addresses, and all student identification numbers were stripped from each transcripts. The transcripts—identifiable by only the transcript study identification number—were then coded.

Table 4.3.5-A.—Institutional characteristics of the postsecondary institutions reported by NELS:88 fourth follow-up study respondents: 2000

Institutional characteristics	Number	Percent
Total*	3,213	100
Institutional control		
Public	1,516	47.2
Private non-profit	984	30.6
Private for-profit	713	22.2
Institutional level		
Four or more years	1,496	46.6
At least two but less than four years	1,229	38.3
Less than two years	488	15.2

* The total included 76 Historically Black Colleges and Universities (HBCU) and 6 American Indian Tribal Colleges.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

From the starting point of the IPEDS-IC institutional database, these schools and their mailing addresses were loaded into a computerized transcript receipt control system developed for the study. The names and addresses of two- and four-year academic institutions were then reviewed by personnel at the American Association of Collegiate Registrars and Admissions Officers (AACRAO), who updated the addresses, provided contact information for registrars and other school officials, and identified closed institutions. JBL and Associates provided similar information for proprietary and less than two-year institutions.[75]

Data Collection Procedures. After updating the addresses in the transcript control system, data collection for the study began. Requests for transcripts were forwarded by Federal Express package to the registrars or other contacts at the schools. These packages included materials that explained the study and provided instructions for returning transcripts and catalogs. Each package contained:

- Lead letters from NCES' Associate Commissioner and the study project director encouraging schools' participation in the study; supporting materials from professional associations and accrediting organizations that endorsed the study (see table 4.3.5-B for a list of these groups);
- A list of students for whom transcripts were requested, including these students' dates of birth, Social Security numbers, and self-reported degrees earned (if any) and periods of enrollment;

[75] These two organizations also provided "refusal conversion" support for these schools, contributing to the study's very low refusal rate.

Table 4.3.5-B.—Organizations endorsing the NELS:88 postsecondary education transcript study: 2000

American Association of Collegiate Registrars and Admissions Officers (AACRAO)
Accrediting Bureau of Health Education Schools (ABHES)
Accrediting Commission of Career Schools and Colleges of Technology (ACCSCT)
Accrediting Council for Independent Colleges and Schools (ACICS)
American Association of Community Colleges (AACC)
American Association of State Colleges and Universities (AASCU)
American Council on Education (ACE)
Association of Catholic Colleges and Universities (ACCU)
Association of Jesuit Colleges and Universities (AJCU)
Council on Occupational Education (COE)
National Accrediting Commission of Cosmetology Arts and Sciences (NACCAS)
National Association of College and University Business Officers (NACUBO
National Association of Student Financial Aid Administrators (NASFAA)
National Association of Independent Colleges and Universities (NAICU)
National League for Nursing Accrediting Commission (NLNAC)

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

- A voucher for the reimbursement of expenses incurred with the request (e.g., transcript processing fees, costs for duplicating out-of-print catalogs);[76]
- A description of *Family Educational Rights and Privacy Act of 1974* (FERPA) requirements authorizing the institutions to release the transcripts without active student consent; and
- Federal Express envelopes and mailing labels for the return of transcripts and catalogs.

Packages returned as undeliverable were traced by project staff. These materials were then remailed to updated contacts and addresses or the schools were recorded as closed. Telephone follow-up with the nonresponding institutions took place two weeks after transmission of the package, with the early follow-up ensuring that the package was sent to the correct person and that this person had received the transcript request.[77] Specially trained institutional contacting staff continued these follow-up calls with nonresponding schools until materials were returned. Over the course of the data collection period, 1,505 institutions (46.8 percent of the schools, overall) received some type of follow-up prompting, with many requiring multiple contacts.

[76] A total of 547 schools returned vouchers. The average voucher was $29, or slightly more than $1 per returned transcript.

[77] New packages were distributed in approximately one-third of these follow-up calls.

School-level data collection results. Table 4.3.5-C provides data collection results for the 3,213 postsecondary institutions reported by the fourth follow-up study respondents. Overall, 3,027 or 94.2 percent of the schools returned transcripts and catalogs.

Table 4.3.5-C.—Institutional characteristics and data collection status of the postsecondary institutions attended by NELS:88 fourth follow-up study respondents: 2000

Institutional characteristic	Sample	Responded		Refused[1]		Closed[2]	
		N	%	N	%	N	%
Total	3,213	3,027	94.2	49	1.5	137	4.3
Institutional control							
Total	3,213	3,027	100	49	100	137	100
Public	1,516	1,498	49.5	12	24.5	6	4.4
Private non-profit	984	957	31.6	19	38.8	8	5.8
Private for-profit	713	572	18.9	18	36.7	123	89.8
Institutional level							
Total	3,213	3028	100	49	100	137	100
Four or more years	1,496	1,479	48.9	18	36.7	0	0
At least two but less than four years	1,229	1,174	38.8	16	32.7	39	28.5
Less than two years	488	375	12.4	15	30.6	98	71.5

[1] Four schools refused to participate; another 45 schools were classified as pending (or passive) refusals at the end of data collection after repeated telephone prompting. Refusing schools accounted for only 130 student transcripts (< 1 percent).

[2] Project staff assigned this code only after determining that the transcripts were not available from another source. Some transcripts were collected from state-level education agencies or other for-profit institutions operated by the owner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Four postsecondary institutions refused to participate in the study. Each of these schools cited confidentiality or legal concerns (as opposed to administrative burden) as their reason for refusing. These schools received subsequent follow-up calls from data collection supervisors and the PETS project director. Each school then received prompting calls from AACRAO or JBL and Associates.[78] An additional 45 schools were coded as passive refusals at the end of data collection. These schools had received multiple calls from both data collection and project staff; they were considered "passive refusals" in the final results.

The possible enrollment period for NELS:88/2000 study respondents spanned an eight-year period, from the end of data collection for the second follow-up study in 1992 to the end of data collection for the fourth follow-up in 2000. Over this period, a number of postsecondary institutions had closed, merged with other schools, changes names, and relocated. Many of these schools (or their successors)

[78] This phased refusal conversion approach was highly effective for most refusing schools. Of the 16 schools that initially refused to participate, 12 schools were ultimately persuaded to provide student transcripts. Four schools could not be convinced to participate.

were located and their student transcripts collected. However, 137 schools received final "school closed" codes from project staff. These schools were initially identified by their failure to return materials. Any school that did not respond and was unlocatable by mail and telephone was assigned for review by project supervisors. In all cases, staff attempted to find current documentation that the school had closed either temporarily or permanently.[79]

Transcript-level data collection results. At the end of data collection for the transcript study, 3,027 postsecondary institutions (or 94.2 percent of institutions reported by NELS:88/2000 respondents) had returned materials. These institutions accounted for a total of 16,020 transcripts (table 4.3.5-D provides the data collection results for these transcripts). Overall 97.9 percent of the transcripts requested for the NELS:88/2000 respondents were returned or otherwise accounted for by the postsecondary institutions.[80]

Table 4.3.5-D.—Transcript-level data collection results for the NELS:88 postsecondary education transcript study: 2000

Transcript status	Number	Percent
Total	16,020	100
Transcript returned	14,654	91.5
No record student ever attended school	858	5.4
Student identified, but no transcript available	179	1.1
Total transcripts unavailable	329	2.1
Problem with transcript	8	0.1
Data collection period expired	27	0.2
School closed	164	1.0
School refused to participate	130	0.8

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

4.4 Further Information on NELS:88 Data Collection Methods and Results

More information about data collection methods and results can be found in the various user's manuals and methodology reports listed at the end of chapter 1.

[79] Approximately one-third of the closed schools had been sold by their original owners and were no longer operational. An additional one-third of the schools had been closed by state or local government or licensing agencies due to financial problems.

[80] This figure includes 15,691 transcripts: 14,654 complete transcripts, 858 cases where the school had no record of the student attending, and 179 where the school had no transcript for the student (e.g., no credit earned, transcript lost or not retained).

Chapter V
Data Preparation and Processing

Data preparation activities spanned each wave of NELS:88, beginning with tracing and securing school cooperation, continuing through monitoring and machine editing, and ending with the preparation of public-use data files and an electronic codebook (ECB). This chapter describes the procedures used to control, prepare, and process NELS:88 cognitive test, questionnaire, and archival records (such as high school transcript) data. Procedures were generally consistent across waves and components, although some differences in data processing stemmed from the different modes of data capture used at different times. Specifically, the three in-school rounds (the base-year and first two follow-up studies) have pronounced commonalties, in that student test and questionnaire data were collected in group administrations, optically scanned, and preserved on microfilm; whereas the two out-of-school rounds for the third and fourth follow-up studies, owing to the post-high school dispersal of the sample, were conducted as individual administrations, typically through computer-assisted telephone interviewing (CATI). In addition, certain changes in naming conventions for numerical codes (also known as consistency codes or reserved codes) reflect the evolution of data processing and ECB software over time, as well as shifts in the topical content of the study. For these reasons, the set of reserved codes used for the third and fourth follow-up studies were more expansive and differ somewhat from the National Longitudinal Study of 1972 (NLS-72) and High School and Beyond (HS&B), and earlier NELS:88 waves. Because of the differences between student survey data collection approaches in the base-year through second follow-up studies, and data capture in the third and fourth follow-ups, this chapter discusses the three in-school rounds together, before proceeding to document the predominantly CATI rounds. The final section of the chapter describes the confidentiality analyses conducted on the base-year through second follow-up data files in order to avoid possible disclosure of school or respondent identities.

5.1 Base-Year through Second Follow-up Editing and Retrieval

For student and dropout questionnaires (including the new student supplement), the first data control and preparation activity was editing questionnaires and retrieving missing information. Interviewers conducted on-site editing of the student and dropout questionnaires at the school or other data collection site, giving special attention to the respondents' answers for all critical items. (Critical items are denoted on the facsimile questionnaires [viewable on NCES' NELS:88 Web Site] by a special retrieval oval; they are also listed in the various past user's manuals.) If the response to one or more of the critical items was missing or indecipherable or had multiple categories marked when only one response was permitted, the interviewer privately pointed out the problem to the respondent. If the sample member indicated that he or she had chosen not to answer the question, the interviewer marked a "no retrieval" response for the item. The "no retrieval" responses were later used during the machine-editing process to assign a "refused" response to the critical items.

Critical items were also designated within the parent, school, and teacher questionnaires, and for high school transcript documents, as well. For these questionnaires and records, editing did not take place in the schools, but rather, in the contractor's offices. Additional follow-up for retrieval of missing or ambiguous critical items took place by telephone.

5.2 Base-Year through Second Follow-up Monitoring and Receipt Control

Once the questionnaires, cognitive tests, and new student supplements were collected, each student and dropout questionnaire was reviewed for completeness and to confirm that the ID numbers were correct. A final disposition code was assigned to each student and dropout indicating whether test data, questionnaire data, or a combination of the two was completed by the sample member. These outcomes were recorded in a microcomputer-based Survey Management System (SMS). Similar review and receipt control procedures were applied to the parent, teacher, and school administrator questionnaires, as well as to transcripts, all of which were subject to telephone retrieval of missing critical items.

5.3 Base-Year through Second Follow-up In-House Editing and Coding

For the optically scanned student and parent questionnaires, the next step was to edit the confidential locator pages for legibility and remove the pages from the questionnaire. In the base-year, student-supplied information on parental occupation was coded by professional coding staff assigned to the research project. In the second follow-up (1992) student questionnaire, respondents were asked to provide the names and locations of the two postsecondary institutions they were most likely to attend after high school. This information was coded using the standard Integrated Postsecondary Education Data System (IPEDS) codes. (IPEDS codes are available only on the restricted-use files.)

A much more intensive coding effort was required for the high school transcripts collected in the fall of 1992 and spring of 1993. The student- and course-level transcript data were entered using a computer-assisted data entry system. The system consisted of sequential data entry screens requesting specific student- and course-level data, such as Scholastic Aptitude Tests (SAT) scores, course title, and credits earned. Screens were grouped by data type (student or course) to facilitate accurate and expedient abstraction. Identifying information (sample member name and identification number and school name and unique school identification numbers) were preloaded into the data entry system. Valid ranges, data field size, and data type (e.g., numeric or text) were specified for each data element; clerks were required to re-enter data failing these checks. Course data were entered with the help of a transcript coding system. This system consisted of a relational database with a Classification of Secondary School Courses (CSSC) look-up table function.

5.4 Base-Year through Second Follow-up Data Capture and Archival Storage

Two data capture methods were employed in the 1988-92 rounds: optical scanning and computer-assisted data entry (CADE). Optical scanning was employed when instruments were collected on a large scale; data entry was utilized when sample sizes were comparatively small. Data capture methods for the base year through the second follow-up are summarized below in table 5.4. All optically scanned instruments were photographed onto microfilm for archival storage. Hard copy questionnaires that had been data entered were not microfilmed but were also archived.

Table 5.4.—Data capture methods used for the NELS:88 base-year through second follow-up studies (in-school waves): 1988-1992

Optical Scanning	Computer-assisted data entry (CADE)
1988, 1990, 1992 Cognitive Tests	1988, 1990, 1992 School Administrator Questionnaires
1988, 1990, 1992 Student Questionnaires	1988 Teacher Questionnaire
1988, 1992 Parent Questionnaires	1990, 1992 Dropout Questionnaires
1990, 1992 Teacher Questionnaires	1990, 1992 New Student Supplement
	High School Transcripts

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

5.5 Base-Year through Second Follow-up Data Processing of Tests and Questionnaires

In each round of the study, data processing activities began with sample selection and continued through receipt control, machine edit, and the preparation of public and privileged use data files and user documentation. Data processing activities varied little among the base year, first follow-up and second follow-up. This section describes the post-processing that was carried out to prepare the data for final release.

5.5.1 Base-Year through Second Follow-up Machine Editing

Conventions for editing, coding, error resolution, and documentation adhered as closely as possible to the procedures and standards previously established for HS&B and NLS-72.

Detection of out-of-range codes was completed during scanning or data entry for all questions except those permitting an open-ended response. The scanning contractor converted the student data to machine-readable form and supplied a raw data tape to NORC. Because of their small number, the new student supplements were not scanned, but were keyed by data entry personnel After receipt of all scanned and keyed data, sequenced machine editing and visual inspection of the output began. The tasks performed included resolving inconsistencies between filter and dependent questions, supplying the appropriate missing data codes for questions left blank, detecting illegal codes and converting them to missing data codes, and investigating inconsistencies or contradictions in the data. Frequencies and crosstabulations for each variable were inspected before and after these steps to verify the accuracy and appropriateness of the machine editing processes.

Inconsistencies between filter and dependent questions were resolved in the machine-editing process. In most instances, dependent questions that conflicted with the skip instructions of a filter question contained data that, although possibly valid, were superfluous. For instance, respondents sometimes indicated "no" to a filter question and then continued to answer "no" to subsequent dependent items. When a filter question indicated that a subsequent question(s) should have been skipped, the dependent questions were set to the value "legitimate skip," with one exception: if the dependent questions were answered in a manner that was inconsistent with the filter but consistent across the dependent items, the filter was back-edited (changed) to agree with the dependent responses. If multiple responses or no answer was given to a filter question, the question was assigned the appropriate reserved

code (see below) and all subsequent questions that might have been skipped were processed as if the respondent had left them blank.

The frequency with which responses were recoded to "legitimate skip" for each skip pattern was closely monitored. Frequency distributions of responses before and after editing were inspected. All filter questions and their respective dependent items were displayed in crosstabulations so that staff could verify the accuracy of the recoding.

After improperly answered questions were converted to blanks, the data were passed through a second step in the editing program that supplied the appropriate reserved codes for blank questions. Where a value was not provided by the respondent, a reserved code filled the field. The reserved codes for these studies[81] and their meanings are as follows:

6 MULTIPLE RESPONSE
7 REFUSED[82]
8 MISSING
9 LEGITIMATE SKIP/NOT IN WAVE.

When the legitimate response of a variable filled more than one column of space, the right-hand column contained one of the above codes and the remainder of the columns were filled with 9's. (In other words, a five-byte missing response would be coded as 99998.)

Critical items for the study—items essential to data analyses or required to locate respondents in the future—followed a somewhat different machine-editing process. Data collection procedures instructed field interviewers to mark the retrieval oval beside each critical item in the questionnaire if an attempt was made to retrieve missing or invalid data from a respondent. The edit program then used these fields to set corresponding blank data to "refused." Since their purpose was to determine the correct reserved codes, retrieval variables did not appear on the final data file. If a critical item was left blank, was not a legitimate skip, and an attempt was made to retrieve the missing data, the item was coded as "8" (missing). If a filter was coded "7" (refused), all subsequent questions that might have been skipped were processed as if the respondent had left each item blank. Filters that were coded "6" (multiple response) or "8" (missing) were handled in the same manner. Items with unusually high nonresponse or multiple responses were checked by verifying the data on the microfilmed questionnaire.

In preparing the public- and restricted-use ECBs for NELS:88/2000 (N0P and N0R), every attempt was made to retain consistency codes, variable labels, and value labels from the earlier data releases for NELS:88 (the N2P/N4P and N2R/N4R ECBs released in 1996). However, NCES standards for developing ECBs changed between the third and fourth follow-up releases making one change mandatory. Prior waves of data included blanks to represent sample members who were not in the wave at the time of the data collection (e.g., sample members freshened into the study in the second follow-up study would not be in the wave for the first follow-up study items and these cases would be blank). Since data collection standards in 2000 do not allow blank data, these cases were combined with legitimate skips (consistency code = 9) in the ECB.[83] Thus, the label for these consistency codes became "Legitimate skip/Not in wave."

[81] Reserved codes for the fourth follow-up study are listed in Section 5.7 of this chapter. Where possible, consistency codes from earlier waves of data were not modified and they should agree with data already been released to the public.

[82] This code was used only when a critical item was missing and the retrieval oval was checked by the field interviewer, indicating that the respondent refused to answer.

[83] There were no remaining unused consistency codes to assign to these blanks, forcing this merging of codes.

Analysts who wish to differentiate between these two categories can examine the gate questions surrounding the legitimate skips. Alternately, data users can also consider the analysis weights or flags for the population. Members of the NELS:88 population who were not in the wave at the time of the data collection in question (e.g., the base-year study) will have weights of "0" or appropriate values on their flags.

5.5.2. Base-Year through Second Follow-up Data File Preparation

The conventions used to assign SAS and SPSSX variable names were as consistent as possible with NLS-72 and HS&B. In the two predecessor studies, variable names were assigned according to the survey wave and the question number. A similar system was developed for NELS:88. For example, BYS56A is from the base-year student survey, question 56, part A. Likewise, F1S7D is from the first follow-up student survey, question 7, part D, while F2S84C is from the second follow-up student survey, question 84, part C. BYP80 is question 80 from the base-year parent survey, F1D11 is question 11 from the first follow-up dropout survey, F2P51C is part C of question 51 from the second follow-up parent survey, and so on.

Constructed or derived variables—including statistical weights, special indicators or flags, and variables that are composites derived from multiple sources—were added to the files in order to promote more accurate analyses of the NELS:88 data. Certain items add information from study sources that would otherwise be unavailable to users; some items reference respondent properties to external standards that would be expensive for individual analysts to create; and other items are recodes or combinations of internal questionnaire sources. A number of derived variables have been created as a convenience for the analyst, rather than as a source of wholly new information.

Generally, the names of the base-year flags, variables, and weights begin with BY; the first follow-up flags and weights begin with F1; and the second follow-up names begin with F2.[84] For school-level variables placed on the student file, the derived variable name begins with the modal grade for the particular round, for example, G8 (for grade 8 in base year), G10 (for grade 10 in the first follow-up) or G12 (for grade 12 in the second follow-up). A few derived variables that were built in the base year do not begin with the prefix "BY." These are SEX, RACE, HISP, API, HEARIMP, HANPAST, BIRTHMO, BIRTHYR. Statistical weights include the string "WT" in the variable name. The case selection flags that must be used in tandem with statistical weights ordinarily have the same variable name stem. For example, the selection flag for use with F2QWT is F2QFLG.

Over the course of the survey, even basic demographics such as sex and race-ethnicity of the respondent were re-examined and updated when and if new or more accurate information became available (thus there is an F1SEX variable on the first follow-up files, and F2SEX on the second follow-up files). The only reserved code used for derived variables was for missing data (8 or "8" in the rightmost column for multidigit variables, e.g., 998).

Final user products for the in-school rounds were public-use and restricted-use files supported by an ECB system. Use of the 1988-92 data in the current re-release of NELS:88 in conjunction with 2000 data is described in the *ECB User's Guide* that resides with the ECB on the CD-ROM and also on the NELS:88 section of the NCES Web Site.

[84] These naming conventions were maintained for the fourth follow-up study, which used the prefix F4. Notably, however, the third follow-up data did not utilize this convention and did not use a prefix.

5.6 Third Follow-up Study Data Control, Preparation, and Processing

Because the 1994 round primarily involved a telephone interview that was administered in a CATI format, post-processing activities were much reduced compared to the earlier rounds. CATI allows many consistency checks to take place during the interview, with the result that extensive post-data collection consistency editing is not necessary.

A large number of composite or derived variables were created in the third follow-up and added to the data files. These constructed variables included measures of high school graduation status, labor force experience, postsecondary education, and family formation. Historical demographic composites were again updated (F3SEX, F3RACE).[85]

Because of changes in ECB software standards, the third follow-up survey deviated somewhat from the reserved code conventions of NLS-72, HS&B, and the earlier rounds of NELS:88, and used negative values as codes for missing data. These codes are

–2 CURRENTLY ATTENDING (assigned when an ending date, e.g., date of school attendance, is asked)

–3 NOT ASKED IN SAQ (hardcopy self-administered questionnaire)

–4 UNCODABLE VERBATIM

–5 NOT APPLICABLE

–6 MISSING

–7 REFUSED

–8 DON'T KNOW

–9 LEGITIMATE SKIP/F3 NONRESPONDENT

–10 INSTITUTION NOT IN 1993/1994 IPEDS FILE

–11 MILITARY TRAINING

–12 FOREIGN INSTITUTION

Just as in the NELS:88 base year through second follow-up, an ECB in both public- and restricted-use versions was produced. However, an additional product was released as well: a table generator called the NELS:88 Data Analysis System (DAS).

As noted above, during the development process for the NELS:88/2000 ECBs (N0P/N0R), every attempt was made to retain consistency codes, variable labels, and value labels from the earlier data releases for NELS:88 (i.e., N2P/N4P and N2R/N4R released in 1996). However, NCES standards for developing ECBs that changed between the third and fourth follow-up releases made one change mandatory. Thus, 93 cases who were F3 nonrespondents (but who were F4 respondents) were combined with –9 (legitimate skips). The label for this consistency code became "Legitimate skip/F3 nonrespondent."

Analysts who wish to differentiate between these two categories can examine the gate questions surrounding the legitimate skips. Alternately, data users can also consider the analysis weights for the

[85] Truncated labels for the third follow-up derived variables released in the NELS:88/94 ECB were abbreviated in the current release (NELS:88 fourth follow-up) ECBs (N0P, N0R).

population. Members of the NELS:88 population who were not in the wave at the time of the data collection in question (e.g., the base year) will have weights of zero. Similarly, cases with MODE > 0 (MODE is the mode of data collection for the third follow-up study) were respondents to the third follow-up study.

5.7 Fourth Follow-up Study Data Control, Preparation, and Processing

Data collection for the fourth follow-up study was conducted almost exclusively with computer assisted interviewing, primarily by telephone (CATI). However, in-person field interviews were also completed with this technology. [86] Field interviewers used the same computer-assisted interview and on-line coding software as the study's telephone interviewers, but on a laptop computer-based platform. Thus, all of the entry of interview data was accomplished by the NELS:88 fourth follow-up CATI-CAPI system. [87] As the interviewers entered the number of the response option selected by the sample members, this number was immediately written to the data file. (The field interviewers data files were downloaded nightly.) Notably, however, some additional coding was required by the telephone and field interviewers. When the respondent selected an "other" response, the interviewer entered text into a verbatim-specify field, which appeared on the screen. Project staff later "coded up" these responses, where appropriate, into existing categories. The remaining "other specify" responses can be located on the study's restricted-use ECB.

The ranges of most items in the CATI and CAPI survey were determined by the codes for the available responses. Additional ranges of responses were also established for continuous measures (e.g., typical income level, years of experience) and these bounds were incorporated into the computer-assisted interview. Where appropriate, these bounds are discussed in the descriptive windows for the fourth follow-up variables. Additional detail is also available in the CATI/CAPI interview facsimile. This facsimile, which shows all skip patterns, wording changes, and fills for the fourth follow-up interview, is located in the Instrument folder on the NELS:88 fourth follow-up ECBs (N0P, N0R) .

5.7.1 On-Line Coding of Responses

During each fourth follow-up telephone or field interview, the interviewer collaborated with the NELS:88 sample member to assign codes to literal responses in three areas: postsecondary institution identification, major/field of study, and industry/occupation.[88] The interviewers entered verbatim responses from sample members, employed a computer-assisted on-line system to identify possible codes for the verbatim strings, and then confirmed the findings with the sample members. The interviewers worked more closely with the sample members for strings that were difficult to code, Each coding operation was subjected to quality control review and recoding procedures by project staff who were expert coders. This review/revision was accomplished on a periodic basis, and expert coders provided general notes to interviewers specifying particular problem areas and suggestions for improving coding quality. Additionally, general telephone survey unit and interviewer-specific information on coding discrepancies were produced periodically to monitor the process of the coding activities. All computer-assisted interviewer coding used software developed by NCES to standardize computer-assisted coding across studies and contractors.

[86] Additionally, a total of 27 hard-copy (self-administered) "interviews" were completed by sample members who could not or would not complete a telephone or in-person interview.

[87] Project staff also used the CATI/CAPI system as the data entry program for returned hard-copy self interviews.

[88] Each type of code may have been assigned several times during the interview. For example, industry/occupation codes were assigned for the respondents' current/most recent employment and for the "job desired at age 30." Major/field of study was entered for up to six degrees; IPEDS identifiers were coded for up to eight institutions.

Institutional coding was needed to assign a 6-digit IPEDS school identifier for all schools specified by the respondents during the fourth follow-up interview.[89] The system relied on a look-up table—in other words, a coding dictionary—of institutions constructed from the 1997-98 IPEDS Institutional Characteristics file of postsecondary schools. Collected and updated annually by NCES, IPEDS is a comprehensive system designed to encompass all institutions and educational organizations whose primary purpose is to provide postsecondary education. Other information in the dictionary, including the institution's degree level (e.g., less than 2-year degrees, 2-year degrees, more than 2-year degrees) and organizational control (e.g., public, private, private-for-profit), was also loaded at this time for use with branching and fills during certain interview items.

Field of study and industry/occupation coding was also performed during the interview. These coding activities also used a dictionary of word-code associations. The on-line procedures for these coding operations included the following steps. First, the interviewer keyed the verbatim text provided by the respondent; then, standard descriptors associated with identified codes were displayed for the interviewer; and finally, the interviewer selected a listed standard description and confirmed this descriptor with the sample member.[90] Importantly, the interviewers were instructed during training to focus primarily on collecting and keying the best possible text for a given coding application. In fact, the code assignment was secondary to the responsibility of getting a sufficient and accurate verbatim text response so that post-hoc coding could be implemented by analysts who might desire to employ their own coding structure.

To evaluate the accuracy and effectiveness of the coding operations, all coding was subjected to quality-control recoding. The recoding also offered the opportunity to provide feedback to the interviewing staff about using the software more effectively. Another result from this quality-control effort was to provide information that could be used in refining the software for future studies. Recoding was done for all on-line coding cases that could not be coded by interviewers. Additionally a 10 percent sample of all interviewer-assigned cases was selected for quality-control recoding. Separate variables for the initial codes and the quality-control-recoded items are included in the study NELS:88/2000 ECBs.

5.7.2 Up-Coding Other-Specify Items

Typically, respondents choose "other specify" options when existing response options are incomplete or when the meaning of an item is not clearly understood. "Other specify" may also be selected by the interviewer when it is unclear how a particular response may be appropriately categorized into existing response options. In the NELS:88 fourth follow-up CATI and CAPI interview, project staff reviewed all of these items and, where appropriate, coded up the responses into the existing response categories. Indeterminate and new response categories were not revised.

5.7.3 Consistency Codes

Standards for ECB design and development changed in the six years between the third and fourth follow-up studies. Thus, the consistency code conventions for NELS:88 fourth follow-up in 2000 differ somewhat from the approaches used for the third follow-up study in 1994, as well as the approach used in

[89] IPEDS codes and institutional information for postsecondary institutions identified by the respondents to the third follow-up interview in 1994 were preloaded into the interview before the start of data collection in January 2000.

[90] The codes/descriptors that had the greatest number of word-code associations were displayed. Multiple codes/descriptors were displayed if there was more than one code having the maximum number of associations. If the interviewer was unable to identify the appropriate code with the respondent's assistance, the item was reported as uncodeable.

the earlier waves of the study and in NLS-72 and HS&B. The approach uses 2-byte negative values for all variables. These values and the labels are presented in table 5.7.3.

Table 5.7.3.—Consistency codes used for the NELS:88 fourth follow-up study: 2000

Code	Label
–1	Don't know
–2	Refused
–3	Legitimate skip
–5	Foreign country
–6	Uncodeable
–7	Not reached-partial/abbreviated interview
–8	CATI/CAPI error
–9	Missing

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Just as in the NELS:88 base-year through third follow-up studies, an ECB in both public- and restricted-use versions was produced.

5.8 Confidentiality: Protecting Respondent Identities from Statistical Disclosure

5.8.1. General Strategy for Disclosure Avoidance

Disclosure avoidance in NELS:88 involved two basic procedures for identification of high-risk variables. First, certain data elements may be identified *a priori* as posing disclosure risk. Variables that constitute virtually unique data signatures pointing to given individuals or schools (for example, many continuous variables; detailed secondary and postsecondary coursetaking histories as recorded in the transcript components of NELS:88), extreme outliers that may be associated with publicly known characteristics of an institution or individual, and finer-grained versions of school-level variables that could be linked to universe files, all fall within the category of pre-identifiable high-risk variables. In all rounds of NELS:88, such data elements were suppressed or altered on the public-use files, though generally maintained (including complete transcript data, geocodes, residence zip codes, IPEDS unit identification numbers for postsecondary institutions, and so forth) on the restricted-use files. Specific confidentiality edits imposed on the public-use files include the selective suppression of variables, the recasting of continuous variables in categorical form, top coding and bottom coding of continuous data, and collapsing of categories in discrete data.

Second, other data elements may be identified *a posteriori*, that is, empirically, as posing a disclosure risk. Such data elements require a disclosure analysis to determine what confidentiality edits are necessary to protect respondent identities. Disclosure analysis was a specific requirement of the three NELS:88 in-school rounds (1988-92), given that the primary risk of disclosure is, first, that a school might be identified, and, second, that after school identification, a specific respondent (such as teacher or student) might be identified. Disclosure avoidance requires that potentially revealing school-level

information from the data files be analyzed in conjunction with data available from school universe files. Where school matches permit institutional identifies to be deductively disclosed, further modification of school-level and sometimes student- or teacher-level variables was required.

5.8.2 *A Priori* Confidentiality Edits

The 1988-92 school, teacher, and student public-use files were all subject to *a priori* confidentiality edits, as were the 1988 and 1992 parent files. For the 1994 and 2000 releases, verbatim strings from the CATI interview were suppressed, as well as zip codes, IPEDS codes, and other potentially revealing information. In addition, some files have not been released in public-use form. These include the High School Effectiveness Study (1990-92) files, school and residential zip code linkages to 1990 Decennial Census data, the NELS:88 High School Transcript Component, and the (to be released, in restricted form only) NELS:88 Postsecondary Education Transcript Study files. (However, summary variables from the high school transcript component have been included on the public-use files, as well as special high school transcript weights that facilitate their accurate use.)

5.8.3 School-Level Disclosure Analysis: Matching with Universe Files

Disclosure analyses were conducted for all three in-school rounds of NELS:88 and are described in detail in earlier reports (see, e.g., Ingels, Scott, Rock, Pollack and Rasinski 1994, pp. 113-115; Ingels, Scott and Taylor 1998). The first step in disclosure analysis assessed disclosure risk against universe files containing both public and private schools. A number of variables with disclosure potential were identified, then categorized as closely as possible across the files in preparation for the calculation of a distance metric. The distance between schools—one on the NELS:88 file and the other on the universe file—was measured using a "code distance" metric. With the code distance measure, results of a code change for confidentiality for a particular school could be readily observed.

A number of distance measures were available for each school—the school's distance with itself (between the two files) and the school's distances with other schools on the universe file. For each NELS:88 school used in the analysis, the distance measures associated with the school were rank-ordered. The actual code distance values associated with each school were, for the most part, irrelevant for this analysis. The important measure is the relative ranking of the school's individual distance compared to its distance from other schools.

In each round, some schools were found to be at risk of disclosure, and recoding was implemented to minimize disclosure risk. Based on the assessment of the analytic importance of the matching variables, it was decided to recategorize variables in the following order: number of teachers, total school enrollment, percent White/Black/Hispanic, and percent free lunch. Grade span and urbanicity would only be considered if changes to these other variables did not sufficiently reduce disclosure risk, and when such adjustments were required, the values were set to "missing" rather than changed. Each time, after recoding was performed, the disclosure analysis was repeated, until no NELS:88 schools were found to be at risk for disclosure from the universe file.

While for most purposes, the public-release files will serve the needs of educational researchers, in some cases, information will be needed that is available only on restricted-use files. NELS:88 restricted-use data are available at no charge on a loan basis to individuals or institutions that obtain an approved license agreement from NCES. To request a license agreement, the individual or institution must provide the following information:

- The title of the survey to which access is desired,

- A detailed discussion of the statistical research project that necessitates access to the restricted NCES survey data,
- The name and title of the most senior official having the authority to bind the organization to the provisions of the license agreement,
- The name and title of the principal project officer who will oversee the daily operations,
- The number, name, and title of professional and technical staff who will access the survey database. Each professional or technical staff member with access to the data is required to sign and have notarized an affidavit of nondisclosure, and
- The estimated loan period necessary for access to the NCES survey database.

To obtain further details and a license agreement please write to:

Data Security Officer
Office of the Deputy Commissioner
U.S. Department of Education
National Center for Education Statistics
1990 K Street, NW
Washington, DC 20006

5.9 Sources of Additional Information

For further details of disclosure analyses and list of suppressed or altered variables:

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M., and Rasinski, K.A. (1994). *NELS:88 First Follow-Up Final Technical Report* (NCES 94-632). Washington, DC: National Center for Education Statistics.

Ingels, S.J., Scott, L.A., and Taylor, J.R. (1998*). NELS:88 Base Year Through Second Follow-Up Final Methodology Report* (NCES Working Paper Series, 98-06). Washington, DC: National Center for Education Statistics.

Appendix A: Quick Guide to Using the NELS:88/2000 Data

Appendix A
Quick Guide to Using the NELS:88/2000 Data

The purpose of the "Quick Guide" is to orient potential users of the NELS:88/2000 data to suggested techniques for working with the data files. Special attention will be paid to topics that will help users avoid the most commonly made mistakes in working with NELS:88 data. This guide is meant to serve as an introduction, not a replacement for the *NELS:88 Base-Year to Fourth Follow-up Data File User's Manual*. The first two sections of this guide provide a brief overview of the NELS:88 survey and available data files. The third section provides general instructions on how to get started using the NELS data and an orientation on the software that can be used to manipulate the data. The final section contains a series of frequently asked questions (FAQs) that NCES receives from users. The questions are followed by responses from NCES.

A.1 Introduction to NELS:88

A.1.1 Overview

During the spring term of the 1987-1988 school year, the National Center for Education Statistics (NCES) initiated a national longitudinal study of 8th-grade students attending 1,052 high schools across the United States. A total of 24,599 8th-graders were surveyed in the base year of NELS:88. Many of these same students were re-surveyed in 1990, 1992, 1994, and 2000. Depending on the year, data were also collected from parents, schools, and teachers, and from extant high school and postsecondary transcripts. In addition, cognitive tests were administered during the first three data collection waves in the 8th, 10th, and 12th grades. In total, the NELS:88/2000 data contain extensive information on a large subsample of the original 1988 sample (approximately 25,000) and include five waves of data (12,144 cases) which can be categorized into the following groupings:

- Student data (junior high/middle school and high school data: 1988, 1990, and 1992)
- Dropout data (1990 and 1992)
- Post-high school data (1994 and 2000) collected after scheduled high school graduation
- School administrator data (1988, 1989, 1990, and 1992)
- Teacher data (1988, 1990, and 1992)
- Parent data (1988 and 1992)
- High school transcript data (1989-92), collected in the fall of 1992[91]
- Postsecondary transcript data (1992-2000), collected in the fall of 2000

[91] Academic transcript data were systematically collected for grades 9, 10, 11, and 12 in the NELS:88 second follow-up. However, when school transcripts contained information about 8th grade or earlier course work, this information was preserved in the transcript file. Likewise, some students took one or more college-level courses while still in high school; this information, too, will appear on transcript files, and it is therefore possible for a student to have postsecondary courses recorded for a year prior to 1992.

A.1.2 Major Features

The major features of NELS:88 include the integration of student, dropout, parent, teacher, and school data; the initial concentration on an 8th-grade student cohort with follow-ups at two-year intervals; the inclusion of supplementary components to support analyses of geographically or demographically distinct subgroups (e.g., regions of the country, Hispanics and Asian subgroups, private school students); and the design linkages to previous longitudinal studies and other current studies.

A.1.3 Research Issues that can be Addressed

The longitudinal design of NELS:88 permits the examination of change in young people's lives and the role of schools in promoting growth and positive life outcomes. By design, the basic unit of analysis is the student, with the parental, school, and teacher components providing contextual information. In particular, data from NELS:88 can be used to investigate a multitude of research topics including:

- Completion of high school for those students who drop out of middle or high school;
- Access and choice to postsecondary schools;
- Persistence and completion of postsecondary degrees;
- The relationship between base-year demographic, academic, and family characteristics, and later educational and employment outcomes;
- The relationship between high school course taking and later postsecondary outcomes;
- The high school and postsecondary experiences and academic performance of language minority students;
- Students pursuit of the study of mathematics and science;
- Transitions from high school and postsecondary education into the world of work;
- Family formation, including marital status and children; and
- Trend analyses with previous longitudinal studies (e.g., NLS-72 and HS&B).

Given the number of issues that can be addressed by NELS:88, it is not surprising that some prospective users of the data have been heard to speculate that the number of potential analyses that can be conducted using the NELS:88/2000 data base is limited only by the imagination of the individual researchers. This view, while not totally defensible, does appear to have face validity when one first examines the numerous files available to analysts. Reality begins to emerge though, once analysts realize that NELS:88 does have both substantive and methodological limitations. There are design constraints (e.g., NELS did not sample regional or area vocational schools) and limitations of the data (e.g., small cell sizes for certain groups of individuals) that must be taken into consideration when planning analyses that use NELS:88.

A.1.4 Must Read Publications

Before a researcher attempts to use the NELS:88 data files, it is strongly suggested that time be spent reading the NELS:88 user's manuals and design documents that reference the NELS:88 base-year and first three follow-up studies. The following list of documents will provide researchers with much of the information that they will need to understand the complexities of the NELS:88 data files.

Manuals/Technical Documentation

- Base-year to Fourth Follow-up Data File User's Manual
  This manual documents the history of NELS:88 from the base-year through the fourth follow-up studies. It is the single most valuable document for working with the NELS:88/2000 data.
- Base-year Sample Design Report
  This report documents the procedures and results of data collection for the NELS:88 base-year survey of 8th-graders.
- Second Follow-up Student Component Data File User's Manual
  This manual documents the data collection activities of the second follow-up data collection and processing activities. It also includes information on the base-year and first follow-up data collections. In addition, it contains a copy of the third follow-up survey instrument.
- Third Follow-up Methodology Report
  This report documents the response rates for the study and the subsampling decisions that were made for the third follow-up study.
- Base-year Through Second Follow-up Psychometric Report
  This reports documents the base-year through second follow-up assessments in mathematics, reading comprehension, science, and social studies/history.

NOTE: Each of these manuals can be found on the NCES Web Site
http://nces.ed.gov/pubsearch/getpubcats.asp?sid=023
From the NCES web site, documents can be searched and downloaded.

NCES Reports

Beginning with the initiation of NELS:88, NCES has produced selected reports using the NELS:88 data. These reports can be found in electronic format on the NCES Web Site under http://nces.ed.gov/pubsearch/getpubcats.asp?sid=023.

Other Reports

To aid researchers in locating reports that have used NELS:88 data, NCES contracted with the Research Triangle Institute (RTI) to produce a comprehensive annotated bibliography of reports (including dissertations) that used NELS:88 data. This bibliography can be found on the NELS:88 Web Site at: http://nces.ed.gov/surveys/nels88/

A.2 Description of NELS:88 Files and Electronic Codebooks

The NELS:88 surveys are available in both public and restricted use versions. For both the public and restricted versions, the following waves of NELS:88 data are included:

1988	→	1990	→	1992	→	1994	→	2000
Base Year (BY)		1st Follow-up (F1)		2nd Follow-up (F2)		3rd Follow-up (F3)		4th Follow-up (F4)

Because of subsampling decisions (especially at the third follow-up study), the NELS:88 data are most efficiently utilized as three separate data sets representing three distinct populations of respondents.

These data sets are discussed below.

A.2.1 Data Set # 1: NELS:88 Base Year (1988) through Second Follow-up (1992)

This file contains 27,394 cases and includes all respondents who participated in any of the first three waves (base year, first follow-up, or second follow-up). This includes base-year ineligibles and freshened respondents (see user's manual for description of base-year ineligibles and freshened students). Data for each student for the period 1988 (base year) through 1992 (second follow-up) can be thought of as one continuous record that contains the following sections:

1. Base-year student questionnaire and assessment data;
2. First follow-up student questionnaire and assessment data;
3. Second follow-up student questionnaire, assessment, and transcript data;
4. First follow-up dropout questionnaire and assessment data;
5. Second follow-up dropout questionnaire, assessment, and transcript data;
6. Base-year school administrator data;
7. First follow-up school administrator data;
8. Second follow-up school administrator data;
9. Base-year parent data;
10. Second follow-up parent data;
11. Base-year teacher data;
12. First follow-up teacher data; and
13. Second follow-up teacher data.

The first 3-4 characters of each variable name identify the section that the variable belongs (e.g., BYS = Base-year Student; F2P = Second Follow-up Parent). At the end of the first and second follow-up student sections, the composite variables (and weights) are followed by responses for freshened students. At the end of the second follow-up freshened student variables, the record contains composite (summary) high school transcript variables.

Restricted-use version: This data set contains 27,805 cases and has not been subjected to the rigorous disclosure risk analysis to which the public-use version employed. (For example, variables have not been top- or bottom-coded.) The restricted version contains the following data sets (or megafiles):

- High school transcript course-level data (714,614 records);
- Base-year through second follow-up school-level data (2,451 records, including links to Common Core of Data, and Quality Education Data universe file; school zip code data; and middle grades school organization and reform practices); and
- Christian School Supplement data (817 records).

A.2.2 Data Set # 2: NELS:88/94 (Base Year through Third Follow-up)

This file contains 14,915 cases subsampled from the base-year through second follow-up file. In addition to the sections listed above for the base-year through second follow-up data set, this file contains third follow-up student data and two separate institution files (institutions attended by NELS:88 students).

It should be noted that **this file does not contain cross-sectional weights for the base-year, first follow-up, or second follow-up files**. The 14,915 cases on this file should not be used for cross-sectional analysis of base-year, first follow-up, or second follow-up data.

Restricted-use version: This data set contains 14,915 cases and has not been subjected to the rigorous disclosure risk analysis to which the public-use version employed. In addition to the student-level data file, the restricted version contains:

- Postsecondary education attendance data (11,560 records)
- Institution information (2,771 records which include the IPEDS code of the institution)

A.2.3 Data Set # 3: NELS:88/2000 (Base Year through Fourth Follow-up)

This file contains 12,144 cases subsampled from the sampling frame for the third follow-up study. In addition to the sections listed above (see BY through third follow-up), this file contains fourth follow-up student data.

Again, ***this file does not contain cross sectional weights for the base-year, first follow-up, second follow-up, or third follow-up files***. The 12,144 cases on this file should not be used for cross-sectional analysis of base-year, first follow-up, second follow-up, or third follow-up data.

Table A-1 provides a summary of the data order and content for the NELS:88 data files by survey year. Note that this table describes both student-level files (including parent, teacher, and administrator data about the student) (items 1-15) and postsecondary institution-level data (items 16-18). The data user should be aware that the institution files may not be identified uniquely by student identification number or may not include student identification numbers. Information about these files and merging the student and school files is described in the fourth follow-up ECB (see especially the descriptions for STU_ID, INCODE, and INSTNUM).

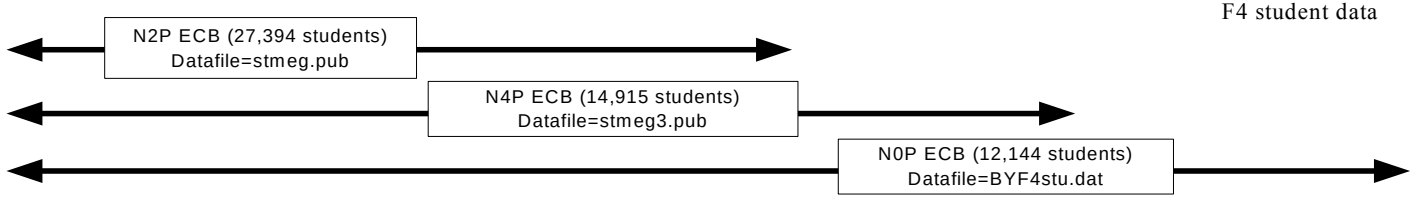
N0P comprises four data files (N0R comprises seven). The four N0P data files are: BYF4STU, PSEF3F4, INSTF3F4, and PSE1994. BYF4STU contains student level information across all rounds (1988-2000), including student, dropout, school, teacher, and parent data, and summary variables from the high school transcript component. PSEF3F4 contains information about postsecondary enrollment in 1994 and 2000. INSTF3F4 contains postsecondary institution-level data for the 1994 and 2000 rounds. Finally, PSE1994 contains data about postsecondary enrollment spells for the entire third follow-up sample (including individuals not represented in 2000).

A.2.4 Description of NELS:88 Electronic Codebooks (ECBs) for Public Release Data

Each of the three public use data sets described above is associated with a separate Electronic Codebook (ECB). The ECBs are tools that allow the user to browse through the lists of NELS:88 variables, variable descriptions, and frequencies.

Table A-1.—NELS:88 student-level datafile content, by survey and ECB: 1988-2000

List Order*	Base Year (1988)	1st Follow-up (1990)	2nd Follow-up (1992)	3rd Follow-up (1994)	4th Follow-up (2000)
1.	BY student data				
2.		F1 student data			
3.			F2 student data		
4.		F1 dropout data			
5.			F2 dropout data		
6.	BY school administrator data				
7.		F1 school administrator data			
8.			F2 school administrator data		
9.	BY parent data				
10.			F2 parent data		
11.	BY teacher data				
12.		F1 teacher data			
13.			F2 teacher data		
14.				F3 student data	
15.					F4 student data



* List order refers to the order the variables appear on the data file and ECB.

Table A-1.—NELS:88 student-level datafile content, by survey and ECB: 1988-2000—continued

List	Base Year	1st Follow-up	2nd Follow-up	3rd Follow-up	4th Follow-up
16.				F3 attendance	F4 attendance
				N0P (16,459 institutions X students) Datafile = pseF3F4.dat	
17.				F3 institution	F4
				N0P (3,217 institutions) Datafile = instF3F4.dat	
18.				F3 enrollment	
				N0P (11,560 students X institutions X number) Datafile = pse1994.dat	

* List order refers to the order the variables appear on the data file.

ECBs allow the user to search a list of variables based on key words or labels; tag (i.e., select) variables for analysis; generate SAS and SPSS syntax for system files; produce printed codebooks of selected variables; import tag files; and access database files for extraction. See the NELS:88/2000 ECB guide for a full description of the functions of the ECB.

Separate ECBs have been developed for the NELS:88 base year through second follow-up, NELS:88/94, and NELS:88/2000 data files. The three ECBs are:

ECB # 1: N2P (NELS:88/92 Public-Use ECB)

This ECB contains the software that will allow researchers to work with the base-year through second follow-up data. The base-year through second follow-up data can be used to examine 8th-graders, 10th-graders, 12th-graders, 8th-graders who make it as far as 10th grade, 8th-graders who make it as far as 12th grade, 10th-graders who make it as far as 12th grade, and dropouts. (27,394 cases – includes 24,599 original base-year students and additional freshened and base-year ineligible students).

ECB # 2: N4P (NELS:88/94 Public-Use ECB)

This ECB contains the software that will allow researchers to work with the base-year through third follow-up data. The base-year to third follow-up data can be used to examine students and dropouts who are still in the sample during the 1994 third follow-up (14,915 cases – subsampled during third follow-up).

Note: It should be pointed out that N4P should not be used to do cross-sectional analyses of 8th-, 10th-, or 12th-graders. The appropriate file for this purpose is the base-year through second follow-up data set (see N2P ECB above).

ECB # 3: N0P (NELS:88/2000 Public-Use ECB)

This ECB contains the software that will allow researchers to work with the base-year through fourth follow-up data. The base-year through fourth follow-up data can be used to examine students and dropouts who are still in the sample during the 2000 fourth follow-up (12,144 cases).

Note: It should be pointed out that this file should not be used to do cross-sectional analyses of 8th-, 10th-, or 12th-graders—the appropriate file for this purpose is the base-year through second follow-up data set (see N2P ECB above).

A.2.5 CD–ROMs

The above NELS:88 data sets, ECBs, and supporting documentation for the NELS:88 base year through second follow-up, NELS:88/94, and NELS:88/2000 data files are located on two separate CD–ROMs.

CD–ROM (NCES 2000-328)[92]

This data product contains:

- NELS:88/92 data, including cross-section data from the base-year, first follow-up, and second follow-up studies;
- NELS:88/94 data (panel data from the third follow-up respondents);

[92] This CD was initially released in March 1996 as NCES 96-128.

- ECB software (discussed above);
- An ECB guidebook; and
- Electronic copies of relevant NELS:88 user's manuals.

CD–ROM (NCES 2002-322)

This newly released product contains the NELS:88 fourth follow-up data (NELS:88/2000) and other useful products. For example, the CD includes:

- Electronic Codebook (ECB) software (discussed in the next section);
- Copy of the ECB guidebook;
- Electronic copy of the *NELS:88 Base-Year to Fourth Follow-up Data File User' Manual;*
- Printable images of the base-year through fourth follow-up study data collection instruments; and
- Annotated bibliography of research employing NELS:88 data.

A.3 Getting Started

This section addresses three questions:

1. What you need to know to get started using NELS:88,
2. How to navigate through the data, and
3. How to generate program syntax to manipulate the data.

These three questions are addressed below.

A.3.1 Question # 1: What do I need to know to get started using NELS:88?

Minimum requirements

1. Obtain a CD-ROM with the NELS:88 data. Two CDs are currently available: NCES 2000-328 includes the cross-sectional and panel data for the base-year and first three follow-ups. NCES 2002-322 includes BY-F4 panel data from the fourth follow-up study.
2. Have access to a computer: The Windows-based ECB requires 4.0 MB of storage space.
3. Develop an analytical strategy for working with data. The sheer number of variables available in NELS:88, and the complex, longitudinal nature of it, make the need for an analytical strategy very important.

Loading and Using the ECB

1. Install the ECB:

   – Place CD–ROM into CD-drive.
   – From Windows, click on "START" and then "RUN."
   – Browse through CD-ROM Drive for "ecbw" folder and open "SETUP.EXE" file.
   – Setup will guide you through the installation of the ECB.
   – Click on ECB icon to run.

2. The user is ready to use the ECB once it is installed. By clicking on each "hot" key on the tool bar found at the top of the ECB screen, the user will quickly understand the structure of the file and the power provided by the ECB to produce data files. At this point, the user should consult the "Electronic Codebook Help Guide" available on the CD–ROM for a specific overview of the ECB functions. (This is a file named HELP.PDF.)

3. Examine the frequencies available for each variable on the ECB. By examining these data descriptions, the NELS:88 user will begin to appreciate the complexity of collecting data from human subjects (legitimate values, legitimate skips, refusals, etc.). It is important to realize that some respondents:

   – Did not respond to the entire instrument;
   – Skipped individual items;
   – Refused to complete selected items;
   – Did not reach the end of the questionnaire;
   – Completed abbreviated versions of the instrument;
   – Made illegal skips; and
   – Responded outside pre-defined valid ranges.

A.3.2 Navigating through the NELS:88 ECB (Identifying a Model and Tagging Variables for Analysis)

1. Define the base population for analysis and whether longitudinal or cross-sectional analysis is required.
2. Develop a conceptual model. What does prior research suggest is happening with the data (e.g., characteristics of students who are likely to drop out of school)?
3. Determine the ***predictor variables*** (e.g., high absenteeism, disadvantaged background, low test scores), ***intervening processes*** (e.g., courses completed, trouble with law, pregnancy), and ***outcomes*** (e.g., event of dropping out, completion of GED, degree attainment, income) that can be used to explain the model.
4. Determine which components (variables) of your model can be addressed with NELS:88 variables. If multiple sources of the same item are available on the data files (e.g., a parent's report of family income versus a student's report), choose the one believed to be most reliable and valid. If the variables that the researcher needs are not available on the NELS:88 files, he/she should consider merging variables from other sources (e.g., Census, Common Core of Data) by working with the restricted-use files (versus the public-use files) through an NCES licensing agreement. A license is necessary to protect respondents since the restricted-use files contain individual level identifiers. Additional information can be located at: http://nces.ed.gov/pubsearch/licenses.asp
5. Re-think original model. If the variables contained on the NELS:88 data files cannot be used to study original model, rethink the model and either modify the model or choose another data set.
6. The user can tag the variables of interest by clicking on the "tag box" next to each variable.
7. The analyst must also remember to choose the appropriate weights and flags for the population of interest. In each data file, flags can be selected to identify a particular part of the population. For example, flags are available to identify whether a student

was a dropout at a particular point in time (e.g., third follow-up). Weights are variables placed on the dataset to compensate for the unequal probabilities of selection and to adjust for non-response. When used with flags, weights allow the analyst to make generalizations about the national populations represented by the various NELS:88 samples. When weights are not used and/or when a flag is used inappropriately, the estimates generated will not be representative of the population.

A.3.3 Generating SAS or SPSS Program Code and Codebook Text

1. After tagging the variables of interest, go to "File" and then "Output."
2. Select the program (e.g., SPSS to generate SPSS program code).
3. Specify directory and name of program code file.
4. Select appropriate button in "Confirmation" box.
5. To view the program code, select "File" and then "View Output."
6. The program code can then be opened in the appropriate software (e.g., SPSS) to generate a working system file and run analyses. It may be necessary to modify the program slightly (check for "execute" statements, period locations and file names). The code should identify the ASCII data file location which will be the CD–ROM.

A.4 Frequently Asked Questions About NELS:88 (FAQs)

Since the first release of NELS:88 data in 1990, NCES staff members have received many questions regarding "proper techniques" for working with the data. In this document, these questions (along with NCES responses) have been categorized into topical areas and presented as a guide. It is hoped that the responses will help users avoid the most commonly made mistakes in working with this important data source. This document is meant to serve as an introduction or supplement, not a replacement, for the Base-Year to Fourth Follow-up Data File User's Manual.

To help the data user identify specific topics of interest, questions and responses have been grouped into the following categories:

FAQs Table of Contents

A.4.1 General and Background Questions

Question: What is NELS:88?

Response: During the spring of 1988, NCES initiated a longitudinal study of 8th-grade students attending 1,052 high schools across the fifty states and the District of Columbia. A subset of these students was resurveyed in 1990, 1992, 1994, and 2000, along with additional individuals who helped to form representative 10th-and 12th-grade cohorts. Data was collected from the students (even after some dropped out of school), their parents, schools, and teachers, and from extant high school and postsecondary transcripts. In addition, achievement tests (assessments in math, reading, science, and social studies) were administered during the first three waves. In total, almost 11,000 pieces of information were collected on a large segment (approximately 12,000) of the original sample of approximately 25,000 students. The NELS:88 survey system includes five waves of data which can be grouped into the following categories:

- 8th-grade/high school student data (1988, 1990, and 1992) including cognitive tests;

- Dropout data (1990 and 1992);
- Post-high school data (1994 and 2000) collected after scheduled high school graduation;
- Junior high/high school administrator data (1988, 1990, and 1992);
- Three waves of teacher data (1988, 1990, and 1992);
- Two waves of parent data (1988 and 1992);
- High school transcript data (1992); and
- Postsecondary transcript data (2000).

Given this breadth of coverage, it is not surprising that some prospective users of the data have been heard to speculate that the number of potential analyses that can be conducted using the NELS:88 database is limited only by the imagination of the individual researchers. This view, while not totally defensible, does have some face validity, especially when data users first examine the numerous files available to analysts. Reality begins to emerge though, once analysts begin to realize that NELS:88 does have some limitations. For example, there are design constraints (e.g., the NELS:88 sample did not sample area vocational schools) and data limitations (e.g., small cell sizes for certain groups of individuals) that must be taken into consideration when planning with the NELS:88 data.

The major features of NELS:88 include the integration of student, dropout, parent, teacher, and school studies; the initial concentration on an 8th-grade student cohort with follow-ups at 2-year intervals; the inclusion of supplementary components to support analyses of demographically distinct subgroups (i.e., Hispanics, Asians, private school students); and the design linkages to previous longitudinal studies and other current studies.

Question: What are some of the terms that I should be familiar with in dealing with NELS:88?

Response: Knowledge of the following terms will help the user in reading through the following questions and responses. Additional information on these and other terms can also be found in appendix E (the NELS:88 glossary) in this data file user's manual.

BYI:	Base Year Ineligible students—students who were determined to be ineligible for the base year many of whom became eligible in subsequent follow-ups
Bias:	respondents differ from nonrespondents
Cohort:	factor in common (year of birth or grade)
Cross-section:	represent events at single point in time
Design effects:	a measure of design efficiency, typically related to the precision of estimates
ECB:	Electronic codebook
Freshening:	adding students to original sample during later waves of data collection to create new grade-representative cohorts
IRT:	Item Response Theory (permits vertical scaling of assessments)

Longitudinal: similar measurements at multiple points in time

Panel: surveying same individuals across time

Weights: used to produce population estimates, or in other words, when one respondent represents a number of others in the population

Question: What are some of the research issues that can be addressed with NELS:88?

Response: The longitudinal design of NELS:88 permits the examination of change in young people's lives and the role of schools, teachers, community, and family in promoting growth and positive outcomes. In particular, data from NELS:88 can be used to investigate a multitude of issues in the context of the family, community, school, and classroom including:

- Students' academic growth over time;
- The transition from 8th grade to high school;
- The process of dropping out of school, as it occurs from 8th through 12th grades;
- The role of the school in helping disadvantaged individuals;
- The school experiences and academic performance of minority students;
- Students' pursuit of the study of mathematics and science;
- The features of effective schools;
- Access to and choice of postsecondary schools;
- Transitions to postsecondary education and the world of work;
- Educational outcomes and their relationships with workplace outcomes;
- Marriage, divorce, and other aspects of family formation among the 8th grade class of 1988; and
- Trend analyses with previous longitudinal studies (e.g., NLS-72 and HS&B).

This question is also addressed more fully in appendix B of the NELS:88/2000 descriptive report (Ingels, Curtin, Kaufman, Alt, and Chen 2002 [NCES 2002-321]).

Question: How were the NELS:88 data collection instruments designed?

Response: Instrument development was guided by the research objectives of NELS:88. Questionnaires and interviews were designed to meet the longitudinal goals of the study; items were chosen based on their utility in predicting or explaining future outcomes as measured in later survey waves. All of the instruments employed in the base-year through fourth follow-up studies were developed to provide continuity and consistency with earlier education longitudinal studies (NLS-72 and HS&B), as well as to address new areas of policy concern and to reflect recent advances in theory. In general, the process for each survey instrument consisted of the following steps:

1. NCES development of list of topics;
2. Contractor development of a content outline;
3. Content outline shared with other government agencies, policy groups, and interested parties;

4. Draft copy of survey instrument developed;
5. Review by the NELS:88 Technical Review Panel (a specially appointed, independent group of substantive, methodological, and technical experts);
6. Survey instrument revised based on reviewer comments;
7. Justification written for components of instruments;
8. NCES review of instruments;
9. Review of instruments by the federal Office of Management and Budget (OMB); and
10. Field testing of instruments, and revision based on field test results.

Scanned images of 15 data collection instruments from all waves of NELS:88 and from all respondent groups are included in the instrument folder on the NELS:88/2000 ECB.

Question: What are the interrelationships among the separate NELS:88 files?

Response: Using common IDs, the individual data files comprising NELS:88 have been merged with each other to form data files containing student, parent, school, and teacher data. By design, the basic unit of analysis for most NELS:88 analyses will be the student. Under this premise, the school administrator, parent, and teacher data can be thought of as providing contextual (e.g., background, school characteristics, "opportunity to learn") data.

Because the base-year of NELS:88 involved the random selection of 1,000 schools from across the United States that contained 8th-graders, the 8th-grade school sample can be used (in conjunction with the 1988 school weight, BYADMWT) as a standalone file in which the school is the basic unit of analysis. The first and second follow-up school files, however, are not nationally representative of high schools with 10th or 12th grades, and therefore no school weight has been generated for them. These schools were not selected by a probabilistic method, but rather, entered NELS:88 by virtue of containing students who participated in NELS:88 during the base-year.

Universe variables have been constructed to provide researchers with a history of the involvement of each student over the base-year and four follow-up studies of NELS:88. These variables show the status of each student during the five data collection periods. For example, a given student may have been eligible for participation in the base-year study, a dropout during first follow-up data collection, back in school during the second follow-up, and subsampled for the third and fourth follow-ups. Universe variables can be used to subset cases to desired populations. Universe variables effectively limit the working data file to respondents who fit selected criteria (e.g., dropouts who are still part of the study at the fourth follow-up). The universe variables can be found at the beginning of the NELS:88 data files.

A special note on the organization and content of the data files in the NELS:88/2000 ECB (N0P or N0R) that relate to postsecondary education institutions might also be useful to data users. In contrast to the student-level files discussed in the previous paragraph, the NELS:88 postsecondary education data files are structured somewhat differently (e.g., at the institution or student-institution levels). Analysts will be able to merge this postsecondary information to the NELS:88 sample members with the careful use of the unique identifiers associated with each file. For example, the PSE attendance data in the NELS:88/2000 ECB (PSEF3F4.DAT) are organized on a sample member (STU_ID) and school identification number (INCODE) basis. INSTF3F4.DAT is an institution-level file with INCODE as the

unique identified. More information on merging these data files can be found in the variable descriptions for the items contained in the ECB.

Question: How is NELS:88 related to prior NCES longitudinal studies?

Response: All of the student and dropout questionnaires employed in the base-year, first follow-up, second follow-up, third follow-up, and fourth follow-up studies of NELS:88/2000 were designed to provide continuity and consistency with earlier education longitudinal studies. (Note: Only student questionnaires were administered in the base-year study; separate student and dropout instruments were administered in the first and second follow-ups; a combined student/dropout interview was employed in the third and fourth follow-up data collection.) Specific items in the NELS:88 documents are replicates of items appearing in instruments from the National Longitudinal Survey of 1972 (NLS-72) and the 1980 and 1982 waves (i.e., the sophomore and senior cohorts) of High School and Beyond (HS&B). The comparability and consistency of items across these three data sets allow for (but are not limited to) the conduct of the following trend analyses:

- NELS:88 1990 sophomores can be compared to HS&B 1980 sophomores;
- NELS:88 1990 sophomores two years later (that is, in 1992) can be compared to HS&B 1980 sophomores two years later in 1982);
- NELS:88 1990 sophomore cohort dropouts (as of 1992) can be compared to HS&B 1980 sophomore cohort dropouts (as of 1982);
- NELS:88 1992 seniors can be compared to HS&B 1980 seniors, and NLS-72 1972 seniors; and
- NELS:88 1992-collected high school transcripts can be compared to HS&B 1982-collected high school transcripts (the 1980 and 1990 sophomore cohorts, two years later).

Comparisons are also possible using high transcript data collected for 1992 (NELS:88 senior cohort members); 1982 (HS&B seniors); and 1987, 1990, 1994, and 1998 high school graduates in NAEP schools. Analyses of postsecondary transcript data are also possible between NELS:88, NLS-72 and the HS&B cohorts.

A.4.2 Sampling

Question: In simple terms, explain how the NELS:88 school and student samples were selected?

Response: Base Year: The NELS:88 schools were selected from a universe file of approximately 40,000 public and private 8th-grade schools across the United States. For the approximately 1,000 public and private schools with 8th grades that were sampled and agreed to participate in NELS:88, complete 8th-grade rosters were produced for each school. From this roster, approximately 24 students were randomly selected. The remaining students on the roster were then grouped by race and ethnicity, and additional 2-3 Asian and Hispanic students were then selected for each school.

First Follow-up: Prior to the first follow-up data collection period, approximately 90 percent of the students moved from a K-8/junior high school/middle school setting to high school. Because of these transitions, students had to be traced to their new schools. In addition, school dropouts needed to be identified, contacted and convinced to participate in the follow-ups. New (freshened) students needed to be added to the sample so that the first follow-up data would be representative of high school

sophomores. Without these freshened students, the file would not be representative of 10th-grade students who did not have the opportunity to participate during the base year (e.g., students who were out of the country in 1988 or who were 1990 10th-graders but not in 8th grade in spring term 1988). Also, 8th-graders who had been considered ineligible (e.g., students with physical or mental disabilities or limited-English proficiency) during the base year but whose eligibility status had since changed were added in. These base year ineligible students were given the opportunity to participate during the 10th grade. Because of the wide dispersal of students, the base-year 8th grade cohort was subsampled.

Second follow-up: Prior to data collection, students needed to be traced. Many stayed in the same high school in which they were surveyed as 10th-graders. Others transferred to new high schools, graduated early, or dropped out of high school. It was also necessary to freshen the sample with 12th graders who did not have the opportunity to be surveyed in prior waves. While no students were subsampled out of the 1992 round, some components (full contextual data from school, transcripts, parent) are based on subsamples.

Third follow-up: Prior to data collection, the decision was made to subsample the NELS:88 respondent population to 14,000 respondents. Selected groups were selected with certainty (that is *all* were selected).

Fourth follow-up: Both respondents and nonrespondents for the third follow-up sample were selected for the 2000 survey. Subsampling was then performed upon sample members who had provide difficult to interview, producing a overall sample of 15,237 members of the NELS:88 population at the start of data collection.

Question: Who do these schools and students represent?

Response: The 1,000 schools represent the approximately 40,000 public and private schools in the United States in 1988 that had 8th-grade students. The nearly 25,000 students sampled in NELS:88 represent the 3,000,000 8th-graders attending schools in 1988, with the exception of Bureau of Indian Affairs (BIA) schools, special schools for students with disabilities, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel serving overseas.

NELS:88 data can be used to represent the following groups:

- 8th-grade schools in 1988 (cross-section): use N2P
- 8th-grade students in 1988 (cross-section): use N2P
- 10th-grade students in 1990 (cross-section): use N2P
- 8th-grade respondents who are still in the study in 1990 (panel): use N2P
- 12th-grade students in 1992 (cross-section): use N2P
- 8th-grade students who participate in 1st and 2nd follow-ups (panel): use N2P
- 10th-grade respondents who are still in study in 1992 (panel): use N2P
- NELS:88 respondents two years after scheduled high school graduation (cross-section): N4P
- 8th-grade students who are still in study in 1994 (panel): use N4P
- 10th- and 12th-grade students who are still in study in 1994 (panel): use N4P
- NELS:88 respondents eight years after scheduled high school graduation (cross-section): N0P

- 8th-grade students 12 years later (panel): use N0P
- 10th-grade students 10 years later (panel): use N0P
- 12th-grade students 8 years later (panel): use N0P

Question: Did the NELS:88 core follow the same group of students through the first, second, third, and fourth follow-up studies?

Response: Although the major objective of NELS:88 was to follow a group of 8th-graders, there were both additions and deletions to the sample as it progressed between 1988 and 2000. The additions included the augmentations of the base-year sample with freshened and Base Year Ineligible (BYI) students. The deletions included subsampling performed several times during the study. These augmentations and deletions are described below.

Freshened students: The NELS:88 sample was freshened with additional 10th-graders in 1990 and additional seniors in 1992. These students were added so that the sample would be nationally representative of sophomores and seniors in those years. Students who were freshened into the sample did not have the opportunity to be selected into the sample during the 8th grade (e.g., they may have been out of the country or out of grade sequence).

BYI students: In addition to freshened students, some base-year ineligible (BYI) students were also added to the first and second follow-up samples. BY students were the individuals (approximately 5%) who were deleted from the 8th-grade sampling process by the school principal or headmaster who determined that these students' lack of English language proficiency, or physical or mental disability, would make it unduly difficult for them to complete self-administered questionnaires or cognitive tests, or would not produce a valid assessment of their abilities. These students were re-evaluated during the first and second follow-up studies. Those whose status had changed (e.g., they had become proficient in English) such that they could now participate were returned to the study.

Subsampling: The NELS:88 sample was subsampled at several points in the study. The initial subsampling occurred during the first follow-up. Because students moved from 1,000 junior high school settings (8th grade) to almost 5,000 high school settings (by 10th grade), a decision was made to subsample those individuals who moved to schools that enrolled few NELS:88 students (i.e., fewer than 10 NELS:88 students). Major subsampling also took place during the third and fourth follow-up study.

In addition, the researcher also should be aware that for budgetary reasons, high school transcripts were not collected from all students (17,100 during second follow-up). Thus, there is a separate weight for students with these transcripts.

Question: Why are there 27,394 cases on the NELS:88 public-use ECB when the base-year sample only contained 24,599 cases?

Response: The base-year to second follow-up data set included in the public-use ECB (N2P) and restricted use ECB (N2R) contain all cases that were ever part of NELS:88. This includes individuals who were ineligible to participate in the base year, as well as those who were freshened into the study during the first or second follow-up studies. It is thus important to use flags and weights to create a working data set to delete cases that may be extraneous to your planned analysis. For example, as illustrated below (from N2R), for the first variable on the base-year 8th-grade student file ("BYS2A"), you might not want to keep the 3,206 cases that were blank for this variable (includes freshened and BYI students) if, for

instance, you are analyzing only base-year data. These freshened and BYI cases were not in the sample during the base year.

BYS2A. Is your mother or female guardian living? (MARK ONE)

Code	Freq	Percent	Label
1	23,967	99.3	YES
2	160	0.7	NO
{blank}	3,206		{NONR/NOT IN SAMPLE THIS WAVE}
8	472		{MISSING}
TOTAL	27,805		

To select cases where there are data for the base year, you can either use the base-year weight "BYQWT" (BYQWT >0) or choose on variable BYS2A (BYS2A NE "blank").

NOTE: For data users of the public-use ECB produced for the NELS:88 fourth follow-up study (N0P), the respondent population contained on this ECB (N = 12,144) reflects the NELS:88 sample members who were retained in 2000 and responded to the fourth follow-up interview. Thus, the frequencies and percentages provided in the example above (excerpted from the N2R ECB [N = 27,805) will not be reproduced by the N0P ECB (which is based upon a subset [12,144] of sample members included in the N2P ECB [N = 27,394]).

A.4.3 Weights

Question: What groups does the NELS:88 data set represent and how do I subset these groups?

Response: The NELS:88 data represent many different populations (e.g., 8th-graders in 1988; sophomores in 1990; seniors in 1992; 8th-graders who were still in school at 12th grade; 8th-graders who dropped out of school by 10th grade; etc.). These groups can be identified through use of flags and analysis weights.[93]

Question: What are these flags and weights?

Response: Flags are variables that were put onto the NELS:88 files to indicate status at a given point in time (e.g., dropout status) or a condition (e.g., handicapped). A series of useful flags (universe variables) can be found at the beginning of the data set. These universe variables give the status of each individual for each data collection (e.g., eligible during base-year, dropout during first follow-up, in-school during second follow-up). Flags can be used by the researcher to select cases for analyses. Most of the flags can be found at the end of each file (e.g., fourth follow-up student/dropout file) with composite and derived variables. For example, F4PNLFL is the base-year through fourth follow-up panel flag that indicates the sample member responded at each of the five waves of NELS:88 data collection.

Weights are variables that are put onto the file to compensate for unequal probabilities of selection and to adjust for the effects of nonresponse. Using weights allows a researcher to make generalizations to the national populations represented by NELS:88. On the NELS:88 student files for the base-year through second follow-up studies (N2P/N2R), there are 12 different analysis weights (these weights are described

[93] *Analysis weights* are also known as *nonresponse-adjusted weights*, and as *final weights*. They are to be distinguished from *raw weights* (or *design weights*), which have not been adjusted to compensate for patterns of nonresponse. Only analysis weights appear on the NELS:88 data files.

below). Each of these weights is specific for a given population. Depending on the group to whom the data are designed to generalize, the individual weights have positive values (>0) for respondents who are members of that particular group and zero (0) for all others. (There are nine different weights for the data presented in the NELS:88/2000 ECBs N0P and N0R.)

Question: Why do we need to use weights with the NELS:88 data?

Response: If we do not use weights, the estimates that we produce will not be representative of the population about which we are attempting to estimate.

Question: Why would unweighted estimates not be representative?

Response: In the base year of NELS:88, approximately 25,000 students were sampled from across the nation. These 25,000 8th-grade students represent the 3,000,000 students who attended 8th grade in the United States in 1988. Thus, each student represents approximately 120 students (3,000,000/25,000=120). But because some policy relevant groups (e.g., Asians, Hispanics, private school students) were over-sampled (greater than their proportion in the population), they are over-represented in the file. Depending on the sampling ratio, the weights for these students would be smaller than the average student. By the same token, other students may represent more than 120 students because they were under-sampled during the study. Nonresponse adjustment must also be taken into consideration because the weights of questionnaire nonrespondents are distributed among the respondents with similar characteristics. Thus, weights reflect both unequal probabilities of sampling and nonresponse adjustments. It is not unusual for a specific weight on a follow-up file to have a range of over 20,000 (e.g., F4QWT ranges from 8.71 to 20,898.71—yes, a single student represents 20,899 other students). Therefore, it is incumbent upon the researcher to use appropriate weighting variables.

WARNING: The researcher should avoid breaking down the sample into such small categories that the analysis is questionable. For example, if a crosstabulation table has a single cell with fewer than 30 cases when the data are not weighted, NCES recommends that the results not be displayed or that the cell be combined with another cell (if appropriate).[94]

Question: Which weights and flags should I use in my analyses?

Response: Table A-4 presents the various weights that can be used with the base-year to fourth follow-up public-use data for NELS:88. Included also in the table are the weights and sample flags that should be used with the analyses; and importantly, the study ECB that contains the data. This includes the fourth follow-up ECB (N0P) and the base-year to second follow-up ECB (N2P), released in 1996. The unweighted sample sizes for the analysis groups and the flags required to subset the data are also included (the required analysis [nonresponse-adjusted] weights for the sample are shown in parenthesis.) (Note: Some of the groups below must be further subset to represent meaningful analysis populations.)

[94] Note that a row can be calculated by determining a mean of 0 and 100 values given to a response where the denominator is the sum of the crosstabulation variable. This frequency can also be shown as a percent of the class variable response by a crosstab response; in this case it is the sum of the row, not a single cell. In this example, no single cell can have a frequency less than 3.

Table A-4.—Descriptions of important NELS:88 analysis groups with unweighted sample sizes and related weights and flags for the groups

Analysis Group	ECB	Number (unwgt)	Flag and weight
1988 8th-graders	N2P	24,599	IF BYQFLG=1; (**BYQWT**)
1990 10th-graders	N2P	17,753	IF F1SEQFLG=0; (**F1QWT**)
1990 cross section (includes dropouts)	N2P	19,394	IF F1STAT=0; (**F1QWT**)
1988 8th-graders in 1990 (panel)	N2P	7,424	IF F1PANFLG=1; (**F1PNLWT**)
1992 12th graders	N2P	16,114	IF F2SEQFLG=0; (**F2QWT**)
1992 cross section (includes dropouts)	N2P	19,220	IF F2STAT=00; (**F2QWT**)
1988 8th-graders in 1992 (panel)	N2P	16,489	IF F2PNLFLG=1; (**F2PNLWT**)
1992 cross sectional analysis of students with high school transcript data	N2P	17,100	IF **F2TRSCWT** > 0
2000 cross sectional analysis with F4 respondents	N0P	12,144	(**F4QWT**)
1988-2000 panel analysis of 8th grade class of 1988	N0P	10,827	IF F4PNLFL=1 (**F4PNLWT**)
2000 cross sectional analysis of students with high school transcript data	N0P	10,310	IF **F4TRSCWT** > 0

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Longitudinal Study of 1988 (NELS:88), 1988-2000.

Question: Thanks for the description of the weights, but what does this mean in practical terms?

Response: Perhaps the following examples can help clarify how the analysis weights can be used to help define your sample:

Example # 1: You are interested in examining the gains in math between the 8th and 12th grades.

Decisions that you need to make:

1. **Which ECB should I use?** In this case, you are examining the BY to F2 period, so you would use N2P. If you wanted to examine the math gains and the impacts this might have had on future earnings, you would want to use N0P, the ECB that contains data from all five waves of NELS:88.

2. **Should I use a cross-sectional or panel weight?** A panel weight would be more appropriate because you are following a group of students over time. In fact, there are two panel weights that would be appropriate (F2PNLWT and F2TRP1WT).

3. **Which of these two panel weights do I use?** You can use either. The choice depends on whether you want to limit your study to those with transcripts or not. If you want to control for math course taking, the logical choice would be F2TRP1WT (the subset of the panel sample who have transcripts).

4. **What do I do with the students who drop out of school?** If you are only interested in students who graduate on schedule, you need to drop high school non-completers from the sample.

5. **What variable do I use to specify cases I want to delete?** The following variable "F2RTROUT" is a composite variable created from high school transcripts showing high school completion status in 1992.

Unweighted

Code	Freq	Percent	Label
01	13,471	77.9	SPRING 1992 GRAD
02	272	1.6	OTHER 1992 GRAD
03	154	0.9	PRE-1992 GRAD
04	22	0.1	DIPLOMA/SPEC ED
05	9	0.1	CERT OF ATTEND
06	420	2.4	STILL ENROLLED IN HIGH SCHOOL
07	2,003	11.6	DROPPED OUT
08	424	2.5	TRANSFERRED
09	25	0.1	AGED OUT
10	4	0.0	DIED
11	10	0.1	LEFT/HEALTH RSN
12	35	0.2	RECEIVED GED
13	87	0.5	OTHER
14	349	2.0	STATUS UNDETERM
{blank}	10,520	.0	{NONR/NOT IN SAMPLE THIS WAVE}

If you just wanted those who were high school graduates, you may want to select those with values of 1, 2, or 3 on **F2RTROUT**. Alternately, your final decision may be to only keep those who met the following criteria:

F2TRP1WT > 0 and (1 LE F2RTROUT LE 3): These include only students with transcript data who graduated from high school some time during 1992 (values of 1,2, or 3 on F2RTROUT). You would use F2TRP1WT to weight this sample.

Example # 2: You are interested in describing the sophomore class of 1990.

Decisions that you need to make:

1. **Which ECB should I use?** In this case, you are examining only the NELS:88/90 data period (i.e., the NELS:88 first follow-up study) so you would use the N2P ECB. Analysts who wish to track the 10th-grade cohort through 2000 must use the N0P ECB.
2. **Do I need a cross-sectional or panel weight?** In this example, a cross-sectional weight would be appropriate because it will also include students who were added (freshened) to the sample during the first follow-up to represent sophomores in 1990. Without these freshened students, the sample would only represent 8th-grade students who made it as far as the 10th grade. It would not have students who did not have the opportunity as 8th-graders of being in the study (e.g., out of the country during base-year sampling in 1988).
2. **What are the cross-sectional weights that I could choose?** The most logical choice is F1QWT. This weight represents a cross-section of respondents in 1990 (both students and dropouts regardless of grade).

3. **How do I limit my analysis to respondents who were 10th-graders in 1990?** There are several flags that can be used to subset your sample. One possibility is to use variable "F1SEQFLG" from the first follow-up. Analysts should also become familiar with the Universe variables—the first variables list on each NELS:88 ECB.

Code	Freq	Percent	Label
0	17,753	84.5	ENROLLED IN 10th GRD
1	726	3.5	ENROLLED IN OTHR GRD
2	2,540	12.1	NOT APPLICABLE
{blank}	6,786	.0	{NONR/NOT IN SAMPLE THIS WAVE}

Thus, for your analyses you would subset for those individuals with F1SEQFLG =0 and F1QWT > 0.

A.4.4 Design Effects

Question: Why do I need to use design effects when I do my significance testing?

Response: Because the NELS:88 sample design involved stratification, disproportionate sampling of certain strata (e.g., oversampling of Hispanics), and clustered (e.g., students within a school) probability sampling, the resulting statistics are more variable than they would have been had they been based on data collected from a simple random sample of the same size. A number of statistical packages (SUDAAN and STRATTAB are two of several possible examples) take account of complex sampling designs in the calculation of standard errors. The AM software does so as well (available for free download: am.air.org).

Question: Is there another procedure that I can use to approximate the correct standard error?

Response: Use of variance estimation software is recommended. However, the data file user's manuals for each wave of NELS:88 include tables of average design effects that can be used to adjust standard errors. Citations for these documents can be found in appendix D of the BY–F4 data file user's manual.

Question: Will you please elaborate on this technique?

NOTE: NCES recommends using a software package that produces adjusted standard errors. The following technique is only an approximation to the correct standard errors.

Response: The following example demonstrates how this technique may be used.

Example: The following question is posed: Do 12th-grade boys spend more time on school days playing video games than 12th-grade girls?

Procedure: Produce a cross-tabulation (using SAS or SPSS) that crosses gender (male vs. female) by self-reported time playing video games.

Step 1—Run this cross-tabulation both weighted and unweighted.

Table A - F2SEX by VIDEO

F2SEX(COMPOSITE SEX) VIDEO - WEIGHTED

Frequency / Percent / Row Pct / Col Pct	DON'T PLAY VIDEO	LESS 1HR PER DAY	1 OR MORE HRS /DAY	Total
MALE	669986	399120	255511	1324616
	25.48	15.18	9.72	50.38
	50.58	30.13	**19.29**	
	40.26	64.66	73.42	
FEMALE	994082	218150	92514	1304746
	37.81	8.30	3.52	49.62
	76.19	16.72	**7.09**	
	59.74	35.34	26.58	
Total	1664068	617270	348025	2629363
	63.29	23.48	13.24	100.00

Table B - F2SEX by VIDEO

F2SEX(COMPOSITE SEX) VIDEO - UNWEIGHTED

Frequency / Percent / Row Pct / Col Pct	DON'T PLAY VIDEO	LESS 1HR PER DAY	1 OR MORE HRS /DAY	Total
MALE	4107	2480	1507	**8094**
	25.10	15.16	9.21	49.47
	50.74	30.64	18.62	
	39.22	64.55	73.62	
FEMALE	6366	1362	540	**8268**
	38.91	8.32	3.30	50.53
	77.00	16.47	6.53	
	60.78	35.45	26.38	
Total	10473	3842	2047	16362
	64.01	23.48	12.51	100.00

Frequency Missing = 2858

NOTE: VIDEO RECODED FROM F2S34A

Step 2—Calculate standard errors for both males and females who play video games more than one hour per school day using the following formula.

St. Error = SQRT [(weighted % * (100 - weighted %))/unweighted n]. Compute Standard Errors for males and females separately.

Male St. Error = SQRT[(19.3 * 80.7)/8094] = **0.44**

Female St. Error = SQRT[(7.1 * 92.9)/8268] = **0.28**

Step 3—For significance testing, pool the standard errors for males and females.

Pooled Standard Error = SQRT[(St. Error for males)^2 + (St. Error for females)^2].

Pooled St. Error = SQRT[$(0.44)^2$ + $(0.28)^2$] = **0.52**

The pooled standard error is used when you are making a comparison between two groups.

Step 4—Correct the pooled standard error for design effect.

Corrected Pooled Standard Error = root design effect (DEFT) X Pooled St. Error

Mean Root Design Effect from table 3.3.1-12, page 55 = **1.89** (Second Follow-up Student Data File User's Manual (NCES 94-374))

Corrected Pooled Standard Error = 1.89 X 0.52 = **0.98**

Step 5—Check for significance.

T-test = (% males - % females)/corrected Pooled St. Error

T = (19.3 - 7.1)/0.98 = **12.44** {significant t test}

Males report playing videos for more than 1 hour almost three times the rate as females (19.3% vs. 7.1%).—See table A for weighted percents and table B for sample sizes (numbers in bold).

SUMMARY – Use weighted percentages (estimates) from table A, unweighted sample sizes from table B, and mean design effect from data file user's manual. See individual user's manuals for more detailed descriptions of design effects.

A.4.5 Electronic Codebooks

Question: Why does the NELS:88/94 CD contain two electronic code books? How are these ECBs different from the ECB on the NELS:88/2000 CD?

Response: The two ECBs on the NELS:88/94 CD (released in 1996) represent two samples. The first, **N2P** (public-use file) or **N2R** (restricted-use file), includes those respondents who participated in the base year, first follow-up, or second follow-up surveys (approximately 27,000 cases). The second ECB, **N4P**

or **N4R**, includes only those cases who were subsampled and responded to the NELS:88 third follow-up study (approximately 14,000 cases). This second set of ECBs (N4P/N4R) has been superseded by the NELS:88/2000 ECB, which expanded data collection to 2000. In other words, N2P/N2R includes the full NELS:88 sample through 1992, or when most of the NELS:88 respondents were completing high school. N0P/N0R follows a subset of these students and dropouts through 2000, when most sample members would have been approximately 26 years old.

N2P = NELS:88 second follow-up public-use file

N2R = NELS:88 second follow-up restricted-use file

N4P = NELS:88 third follow-up public-use file (superseded by N0P)

N4R = NELS:88 third follow-up restricted-use file (superseded by N0R)

N0P = NELS:88 fourth follow-up public-use file

N0R = NELS:88 fourth follow-up restricted-use file

N0T = NELS:88 fourth follow-up postsecondary transcript file (restricted)

N2P or N2R should be used when the investigator is interested in the following populations:

- 1988 8th-graders
- 1990 10th-graders
- 1992 12th-graders
- Base-year to first follow-up panel
- Base-year to second follow-up panel
- First follow-up to second follow-up panel
- Dropouts during first and second follow-ups
- 9th- to 12th-grade transcript studies

N0P or N0R should be used when the investigator is interested in the following populations:

- 1991-92 school year high school graduates 8 years later
- 12th grade to fourth follow-up panel (spring 1992 seniors in 2000)
- 10th grade to fourth follow-up panel (spring 1990 sophomores in 2000)
- Base-year to fourth follow-up panel (spring 1988 8th-graders in 2000)

WARNING: The base-year, first follow-up, and second follow-up cross-sectional weights are not on the N0P or N0R ECBs. If the proposed analysis is examining middle school or secondary school issues, then the researcher should be using N2P or N2R.

Question: What is the sequence of data files on the NELS:88 ECBs?

Response: The files containing the N2P and the N0P data are arranged in a specific order. For example, the NELS:88 N2P ECB begins with a series of five universe variables that describe how the respondent entered the study and the student's status at each wave of data collection (i.e., base-year, first follow-up, and second follow-up studies). These five universe variables are followed on N2P by:

1. Universe status of respondents at times of interviews (1988, 1990, and 1992).
2. Base-year student variables followed by base-year composite variables and weights.
3. First follow-up student variables followed by
   a. first follow-up composite variables and weights; and
   b. first follow-up freshened student responses
4. Second follow-up student variables followed by
   a. second follow-up composite variables and weights;
   b. second follow-up freshened student responses; and
   c. composite transcript variables
5. First follow-up dropout variables followed by first follow-up composite variables and weights.
6. Second follow-up dropout variables second follow-up composite variables and weights.
7. Base-year school variables followed by base-year composite variables and weights.
8. First follow-up school variables followed by first follow-up composite variables.
9. Second follow-up school variables followed by second follow-up composite variables.
10. Base-year parent variables.
11. Second follow-up parent variables.
12. Base-year teacher variables.
13. First follow-up teacher variables.
14. Second follow-up teacher variables.

Since the fourth follow-up data include postsecondary education information presented at institution and student by institution levels, the order of files in N0P data is slightly different:

STUDENT DATA

1. Universe status of individuals at times of interviews (1988, 1990, 1992, 1994, and 2000)
2. Base-year student variables
3. First follow-up student variables
4. Second follow-up student variables
5. First follow-up dropout variables
6. Second follow-up dropout variables
7. Base-year school variables

8. First follow-up school variables
9. Second follow-up school variables
10. Base-year parent variables
11. Second follow-up parent variables
12. Base-year teacher variables
13. First follow-up teacher variables
14. Second follow-up teacher variables
15. F3 student/dropout variables (multiple sections from CATI)
16. F4 student/dropout variables (multiple sections from CATI)
17. F4 derived variables and weights

POSTSECONDARY EDUCATION DATA

18. F3 and F4 student*institution-level postsecondary education information
19. F3 and F4 institution-level postsecondary education information
20. F3 student*institution*episode postsecondary education information

Question: When I receive my NELS:88 CD, what are some of the steps that I should follow to check out my CD?

Response: The following steps may help you get a better understanding of the NELS:88 data.

Step 1—Make sure that you have the right file

A general rule that should be followed by all researchers when they receive data from the government or any other source is to check the file for accuracy. Does this file include what you think it does? The following questions should be answered for the NELS:88 CD.

1. Does the NELS:88/2000 CD contain the files listed in the documentation? Check directory and sub-directories.

2. After running frequencies on selected variables on the data file (e.g., first variable, last variable, and five at random in between first and last), do the frequencies agree with those shown for ECB or user manuals? If not, did you receive the correct version?

3. Do the analysis weights (final nonresponse-adjusted weights) contained on the data files allow you to replicate weighted frequencies found in the user manuals? You may want to run weighted frequencies on a single variable using each of the weights contained on the file.

Step 2—Understanding the Data

Assuming that you performed the above steps and you are confident that the files appear to contain what you hoped they would, it is now time to start learning about the files that you will be working with. Start by asking the following questions:

What were the processes involved in getting data from students via questionnaires and cognitive tests to the medium (CD) that you now possess? Just because you did not collect the data does not mean that you do not need to know the procedures that were involved in collecting and processing the data. You also need to understand the quality control checks that were performed by the contractors in processing the data. It is important to realize that some respondents did not respond to entire instruments; other respondents skipped individual items. For example,

a) some refused to complete selected items,

b) some did not reach the end of the questionnaire,

c) sometimes abbreviated versions of instruments were used in data collection,

d) some respondents made illegal skips, and

e) some respondents responded outside valid ranges.

2. What can I do to further my understanding of the cases and variables that I plan to use? You can perform your own quality control procedures by answering the following questions:

   Are the cases that I selected representative of the population to which I wish to generalize? How do the various breakdowns of the data compare to known population numbers? Is my sample biased—do nonrespondents look different from respondents?

A.4.6 Composite Variables

Question: What are the advantages of using composite variables in my analyses?

Response: Composite variables (also called derived variables) were developed for NCES by NORC and RTI to help the researcher analyze the NELS:88 data. These variables were usually created from two or more variables and are often considered to be more accurate measures of the underlying concept than the individual variables that were used to create them. For example the base-year socioeconomic status variable (BYSES) is a composite variable made up of five separate variables from the base-year parent questionnaire representing both parents' education levels, both parents' occupations, and family income (e.g., BYP30, BYP31, BYP34B, BYP37B, and BYP80).

A.4.7 Model Building

Question: How do I select variables for a working data file?

Response: The following sequence of steps will help you to produce your own working data file.

Model Building

After a researcher understands 1) how the NELS:88 data were collected and processed, 2) limitations of the data, and 3) research issues that can be addressed, he/she is ready to

begin selecting variables for his/her working data files. The working data file will be used to test the models that are derived from previously developed conceptual models. Before a working data set is created though, the following steps are suggested:

1) Develop a conceptual model—What does prior research suggest is happening with the data (e.g., characteristics of students who are likely to drop out of school)?

2) Determine the predictor variables (e.g., high absenteeism, disadvantaged background, low test scores), intervening processes (e.g., courses completed, trouble with law, pregnancy), and outcomes (e.g., event of dropping out, completion of GED) that can be used to explain the model.

3) Determine which components (variables) of your model can be addressed with NELS:88 variables—If there are multiple sources of data (e.g., student, parent, school, teacher) available on the NELS:88 data files, choose the ones that the researcher believes are most reliable and valid (e.g., parent report of family income is more reliable than student report). If the variables that the researcher needs are not available on the NELS:88 files, he/she should consider merging variables from other sources (Census, Common Core of Data) through licensing agreement.

4) Rethink original model—If the variables contained on the NELS:88 data files cannot be used to study original model, rethink the model and either 1) modify the model or 2) choose another data set.

Once the above steps have been completed, it is time to subset the NELS:88 data files into working data files. The following steps are suggested:

1) Determine which variables are needed from each of the NELS:88 data files. For example, the model may specify that the following variables are needed from selected files. For example, base-year student data on aspirations, TV exposure, and hours of homework per week can be used to predict fourth follow-up outcomes like highest education attainment, satisfaction with occupation, or income.

2) Determine the analysis population that you wish to work with. This will need to be known so that correct survey questions, filters, and weights can be tagged and included in the extracted files.

3) Use the ECB (N0P or N2P) to tag variables and then create a SAS or SPSS system file.

3) Check log of computer runs to determine if program is doing what you want it to do rather than the directions provided by computer program.

4) Run frequencies and/or means on all variables in working data file to serve as codebook and documentation.

5) Document all cases that are excluded from the working data files (e.g., who is being deleted from the analysis?).

Question: How do I subset data files?

Response: It is very important for the user of NELS:88 data files to learn the proper techniques for subsetting the data. If a user does not correctly subset the files, there will be extraneous cases on the working data file that can potentially complicate the analyses and result in erroneous findings.

Why subset? The NELS:88 data were designed to serve many different audiences. Because of this role, the data can be subset to represent many different populations (e.g., the 8th-grade class of 1988; the 12th-grade class of 2002; the panel of 8th-graders who participated in the study from the base-year through the four subsequent waves). By applying the analysis weights, the user can specify the population that is to be examined. (Of course, some flags [e.g., F1BYQFLG] can also be used for subsetting.) The following examples demonstrate how the researcher can use weights to subset the NELS:88 data.

> Example # 1: Using the NELS:88 base-year through second follow-up ECB (N2P), the user can specify by using the subsetting statement (in SAS or SPSS) IF BYQWT > 0 that he/she only wants those cases that have valid responses for the base-year survey. Instead of 27,000 cases, the file will contain 24,599.
>
> Example # 2: If a user is interested in postsecondary education outcome variables, he/she will want to use the base-year through fourth follow-up ECB (N0P). By specifying use of this ECB, the user has already limited the number of valid cases to 12,144 (instead of 27,000). Thus 15,000 invalid cases have been eliminated.

NOTE: Analysis (final nonresponse-adjusted) weights such as BYQWT have positive values for cases that should be included in the population of interest and values of "0" for all other cases. Thus, programs that subset the N2P ECB with "IF BYQWT > 0" will only keep cases that have positive values, or in other words, those cases with responses in the base year from eligible sample members.

A.4.8 Transcripts

Question: How do I use the high school transcript file?

Response: The transcript file is a separate large file on the N2R restricted CD-ROM that contains separate records for each course that a student took while attending high school. Each separate course is assigned a course ID (F2RCSSC) that can be used to group courses (e.g., F2RCSSC codes ranging from 270000 to 279999 represent math courses).

Information on postsecondary education transcripts can be found in another ECB.

A.4.9 Privileged or Restricted-Use Data

Question: When do I need to use the restricted-use data file?

Response: You should consider using the NELS:88 restricted files when you need data that are more detailed (e.g., individual transcript course data; characteristics of student's neighborhood) than data contained on the public release files. For example, if you need transcript data or zip code data, you should think about obtaining a NCES license agreement.

Contact Cynthia Barton at (202) 502-7307 or e-mail at Cynthia.Barton@ed.gov for information on how to obtain a license.

Appendix B: NELS:88 Data Not Available on Public-Use ECBs

Appendix B
NELS:88 Data Not Available on Public-Use ECBs

The purpose of this manual is to provide guidance and documentation for use of the NELS:88 data set, and in particular the NELS:88 public-use Electronic Codebooks (ECBs). While for most purposes of most users, the public-use files will provide a sufficient basis for addressing their research questions, for other purposes the public-use files may be insufficient. We therefore summarize here some of the other files that have been created and are available from NCES, generally under special license for restricted use. Many, though not all of these files have been incorporated into the three restricted-use longitudinal ECBs: 1988-1992 base year through second follow-up (N2R), 1988-1994 third follow-up (N4R), and 1988-2000 fourth follow-up (N0R).

The extended NELS:88 database comprises the following supplements and files beyond those to be found on the three public-use ECBs (1988-92, N2P, with an inclusive participating sample [participating any of the three rounds] of 27,394; 1988-1994, N4P, with an exclusive [1994-defined] participating sample of 14,915; and 1988-2000, N0P, with an exclusive [2000-defined] participating sample of 12,144).

1. The **Enhancement Survey of NELS:88 Middle Grades' Practices**, a supplement of base-year school principals, was conducted in the fall of 1989, following the base year.
2. The **Christian Schools Supplement**, focusing on a nationally representative sample of Reformed Christian Schools, was conducted in the base year and second follow-up.
3. The **Early Graduate Supplement file** contains additional data, collected in the second follow-up, for NELS:88 students who graduated (received a high school diploma or a GED) before the spring of 1992. This is a public-use file.
4. The **cognitive test item data files** contain sample members' responses to items on the base year, first follow-up, and second follow-up multiple choice cognitive test batteries.
5. The **high school transcript file** contains detailed course-taking records for all participants in the NELS:88 high school transcript component and is available under restricted-use conditions only. (More limited transcript information—summary variables only [and appropriate transcript weights]—appears on the public-use ECB.)
6. **Unedited versions of the NELS:88 public-use files**. The original NELS:88 microdata, prior to editing for confidentiality purposes, are preserved on the NELS:88 restricted-use ECBs.
7. The **High School Effectiveness Study (HSES) ECB**. School selection probabilities were simulated for a subsample of urban and suburban NELS:88 first follow-up (1990) schools. Additional students were selected within these schools so that the student samples were both increased in size and made representative of the school. The same schools were returned to two years later in the second follow-up (1992). This design was implemented to permit NELS:88 to better study school effects, in a manner analogous to that achieved with the High School and Beyond Sophomore Cohort in 1980-82. HSES data are not included on the regular NELS:88 ECBs, but a special HSES ECB is available from NCES.
8. The **expanded sample file**, containing school and student information for both eligible and ineligible members of the 8^{th}- and 10^{th}- grade cohorts, permits researchers to generate more accurate dropout estimates for the 8^{th}-and 10^{th}-grade cohorts and to explore the magnitude of bias on key estimates associated with student exclusion or ineligibility.

9. The **NELS:88-HS&B 1990/1980 equated math score file** allows comparison of the mathematics performance of 1990 NELS:88 sophomores with the performance of the 1980 HS&B sophomore cohort.
10. The **NELS:88 1990 Census data files** contain selected zip code-level variables from the 1990 Census tapes for the NELS:88 base-year through second follow-up responding school samples. Additional files link 1990 Census zip code data to students' 1988, 1990 and 1992 residential address (this information is not available for the HSES sample).
11. The **NELS:88 QED-CCD-SDDB link files** include variables enabling researchers to link NELS:88 schools to external school and district frames, including the Common Core of Data (CCD), the School District Data Book (SDDB), and the Quality Education Data, Inc. (QED) files.
12. The **NELS:88 QED district and school data files** contain variables characterizing the public districts, Catholic dioceses and schools of all types that participated in the NELS:88 base-year, first follow-up and second follow-up surveys. These files are subsets of the master files provided by Quality Education Data, Inc. (QED) of Denver, Colorado and used in each survey wave for sampling or as a source of contacting information.
13. **NELS:88 Postsecondary Education Transcripts (PETS) data file.** Postsecondary education transcript data were collected by RTI in the latter part of 2000. This file does not appear on the NELS:88/2000 restricted-use ECB (N0R). NCES is currently preparing a NELS:88 postsecondary education transcript file that will be comparable to those produced for NLS-72 and the two cohorts of HS&B.
14. **Other NELS:88/94 and NELS:88/2000 special restricted files.** Certain information collected in the third and fourth follow-ups such as, in particular, many of the variables pertaining to postsecondary institutions, and some of the verbatims collected in telephone interviews, are available only on the restricted-use ECBs.

More details about the 14 data sources listed above are provided immediately below.

B.1 Enhancement Survey of NELS:88 Middle Grades' Practices

The Survey of Middle Grades, Practices enhanced the NELS:88 base-year school questionnaire by collecting new information to monitor middle grades reform in the schools attended by NELS:88 8th-graders. Specifically, the enhancement survey augmented the information in the base-year school administrator questionnaire with additional information on school organization, guidance and advisory periods, rewards and evaluations, curriculum and instructional practices, interdisciplinary teams of teachers, transitions and articulation practices, involvement of parents, and other practices recommended for middle grades reform. The questionnaire for this supplemental survey was designed by the Center for Research on Effective Schooling for Disadvantaged Students (CDS) of the Johns Hopkins University. The survey was funded by the Office of Educational Research and Improvement, U.S. Department of Education, and the data collection was conducted by NORC. The school principals who provided base-year information in the NELS:88 school questionnaire were asked to participate in this enhancement survey between late October 1988 and February 1989. The enhancement questionnaire is reproduced in appendix F of the *NELS:88 Second Follow-Up: School Component Data File User's Manual* (Ingels et al., NCES 94-376) and resides on the NELS:88 instrumentation section of the NCES Web Site.

B.2 Christian Schools Supplement (CSS)

In 1988, a sample of Reformed Christian schools that were members of the Christian Schools International (CSI) Organization was drawn to supplement the NELS:88 base-year school sample. The sample was selected from CSI schools with probability proportional to 8th-grade size. Two disproportionately large school units were double-sampled. Of the initially contacted 58 schools, 41 schools agreed to participate. (Due to the double-sampling of the two schools, the number of sampling units was 43.) The student sample drawn from the selected CSI schools constitutes a nationally representative sample of 8th-graders attending CSI schools in 1988 and supports both cross-sectional and longitudinal analyses. Sampled students and their parents, teachers, and school administrators were surveyed in the spring of 1988, during the NELS:88 base year. Students completed both the cognitive test battery and the student questionnaire during the in-school survey sessions held in their schools. Base-year CSS sample members still enrolled in school, their school administrators, and their parents were surveyed again in the spring of 1992, during the NELS:88 second follow-up. Instruments used in the 1988 and 1992 CSS surveys were identical to those completed in the core NELS:88 base-year and second follow-up surveys. (CSI schools also constitute a separately analyzable sampling stratum within the NCES Schools and Staffing Survey.)

B.3 Early Graduate Supplement

The early graduate supplement to the second follow-up student questionnaire was included for persons who had already completed high school at the time of the second follow-up data collection during the spring of 1992. Specifically, early graduate supplement data are provided for respondents who:

- completed the main portion of the second follow-up student questionnaire;
- answered "Already graduated" to Q. 6A in the main portion of the questionnaire ("What grade are you in?"); and
- answered at least one item in the early graduate supplement (Q.114-Q.127B of the second follow-up student questionnaire).

The NELS:88 supplement paralleled the High School and Beyond (HS&B) early graduate supplement and collected information about when the student graduated, why he or she chose to graduate early and who helped in making the decision and the student's activities since early graduation (continuing his/her education, working, participating in a training program, actively serving in the military, etc.) If the student attended a two- or four-year college or vocational school, additional information was sought about when, where and how often the student attended the school. If the student worked, information about the type and length of employment was requested. The NELS:88 early graduate supplement differs from the HS&B supplement in one respect: NELS:88 included in the early graduate sample members of the NELS:88 cohort who had graduated by alternative means, such as the GED, whereas HS&B did not. (Some NELS:88 dropouts obtained examination certification of high school equivalency prior in time to the modal graduation date of their cohort peers who remained in high school, in effect becoming, despite their dropout status, early graduates.) Early graduates who earned a GED can be separated from those who earned a high school diploma to compare NELS:88 and HS&B early graduates, using responses to NELS:88 second follow-up student questionnaire item F2S6B.

B.4 Base-Year Through Second Follow-up Cognitive Test Item Files

The three cognitive test item files contain raw (unscored) choices selected by test takers in the NELS:88 base year, first follow-up, and second follow-up. In each of the three waves, subsets of test items were selected from an overall pool for each of the four subject areas (reading, mathematics, science,

and history/citizenship/geography) to make up the test forms administered to survey participants in that year. The overlap among the test forms allowed the development of a common score scale that could measure change over time even though participants answered different assortments of test questions at each administration. In the base year, all participants received the same test form. On the basis of their performance in the base-year, students were assigned reading and math tests of different average difficulty in the first follow-up in order to increase accuracy of measurement. Similarly, second follow-up reading and math tests were assigned on the basis of performance in the first follow-up. There were two levels of the reading test and three levels of the math test in each of the latter two years. (In the first and second follow-up surveys, freshened students and prior-round nonrespondents were assigned the low-difficulty reading test and the middle-difficulty math test.)

For further information, including a complete test item layout, see Ingels, Scott, and Taylor, *NELS:88 Base Year Through Second Follow-Up Final Methodology Report*, NCES Working Paper 98-06, appendix A, downloadable from the NCES Web Site.

B.5 NELS:88 High School Transcript Component Files

The public-use ECBs contain only transcript summary variables. The restricted-use ECBs make available the entire transcript component. Two data files, the student file and the course file, constitute the transcript component data set. The student file contains one record for every sample member for whom a transcript was collected. The student record includes the student identification number; transcript survey weight; student-level variables abstracted from transcripts (e.g., number of absences per school year, class rank, class size) flags and composites copied from the base-year, first follow-up, and second follow-up student component data file; and flags and composites constructed from course-level data. The course file contains one record for every secondary school course taken by sample members on the student file. Course records consist of the student identification number; the term and course identification number sequentially assigned to the course by the data entry program, and course-specific variables, including course title, course code (assigned from the Classification of Secondary School Courses), grade level at the time the course was taken, credits earned, and grade. The NELS:88 high school transcript files are comprehensively documented in the *NELS:88 Second Follow-Up Transcript Component Data File User's Manual* (Ingels, Dowd, Taylor, Bartot, Frankel and Pulliam 1995, NCES 95-377).

B.6 Unedited Versions of the NELS:88 Public-Use Files: Restricted-Use ECBs

There is a parallel restricted-use ECB for each of the three public-use ECBs (1988-1992, 1988-1994, and 1988-2000). Restricted-use ECBs require a licensing agreement with NCES. The restricted files differ from the public-use files in several respects. First, sometimes they contain components or information that do not appear on the public files (for example, the high school transcript component, the school-level 1990 Census zipcode variables, link variables to external sources [QED, CCD, SDDB], and so on). Second, sometimes (this is the case only for the NELS:88 1988-1992 ECBs, N2P as contrasted to N2R) they contain additional cases (see appendix I of Ingels, Scott, and Taylor, *NELS:88 Base Year Through Second Follow-Up Final Methodology Report*, NCES Working Paper Series [NCES 98-06 for details]). Third, they sometimes contain variables that were suppressed for confidentiality reasons, or altered (again see appendix I of NCES 98-06 for a list of 1988–1992 suppressed or altered variables).

B.7 HSES Baseline and Follow-up (1990-1992) Data Files and ECB on CD-ROM

In 1995 a CD-ROM was produced for the High School Effectiveness Study, including both the 1990 and 1992 waves of HSES and all HSES components (student—including both multiple choice and constructed response test data, school administrator, teacher, parent, transcript, and course offerings). Like the 1996 release of the base-year to second follow-up NELS:88 data, datasets with the same unit of analysis were combined to create files with multiple records per case. The HSES files are supported by an electronic codebook system included on the CD-ROM. The *NELS:88 High School Effectiveness Study: Data File User's Manual* (Scott, Ingels, Pulliam, Sehra, Taylor, and Jergovic, 1996) provides a complete description of the HSES data files.

B.8 Base-Year Through Second Follow-up Expanded Sample File

The NELS:88 second follow-up expanded sample file was constructed to allow licensed researchers to generate more accurate national dropout rate estimates for the 8th-grade cohort as well as more accurate and HS&B-comparable sophomore cohort dropout statistics. In addition, the file can be used to more fully characterize students who were excluded from the NELS:88 base-year sample categories of students who typically have been excluded from national and state assessments and to explore the biasing impact on estimates for the ideal target population that stem from ineligibility and exclusion rules. Cases on the expanded sample file include the grade 8 and grade 10 cohort members who appear on the NELS:88 core restricted-use files, plus ineligible grade 8 or grade 10 cohort members who have never before appeared on a NELS:88 core restricted-use file, except for the transcript component files. Included in the group of ineligible students appearing on the expanded sample file are base-year ineligible (BYI) students who remained ineligible in the first and second follow-ups of NELS:88 and students who were freshened in the first follow-up but were found to be ineligible and remained ineligible in the second follow-up.

A number of variables have been specifically constructed for use with the expanded sample and are included on the file, including student and school background variables, enrollment and out-of-sequence indicators, a variable indicating reason for ineligibility for the student survey (if applicable), cohort flags and a statistical weight, F2EXPWT, which is the only weight that can be used with the expanded sample. The enrollment status indicators for the expanded sample, F1ENREXP and F2ENREXP, include imputed values for cases with missing enrollment data. Only the variables created specifically for the expanded sample should be used with the sample.

For further information, including a listing of derived variables available on the file, see Ingels, Scott, and Taylor, *NELS:88 Base Year Through Second Follow-Up Final Methodology Report*, NCES Working Paper 98-06, downloadable from the NCES Web Site.

B.9 NELS:88 1990/HS&B 1980 Equated Math Scores

In order to compare the mathematics performance of the 1980 HS&B sophomore cohort with that of the 1990 NELS:88 sophomores, it was necessary to put the 1980 mathematics test scores on the same scale as the 1990 scores. The NELS:88 mathematics test was originally designed to be linked to the HS&B mathematics test scores. This was accomplished by including 16 quantitative comparison items from the HS&B test in the NELS:88 test. The mathematics test was the only cognitive test in the NELS:88 battery that shared sufficient items with its counterpart measure in HS&B to enable a reliable cross-walk between the two scales.

Linking was carried out by estimating the item response theory (IRT) parameters for the common items using the NELS:88 sophomore sample and then putting the remaining non-overlapping HS&B items on that scale. Before the final linking was carried out, the item traces for the common items were estimated separately for the two populations and compared to insure that they were "behaving" similarly in the two populations. A final check on the validity of the equating was carried out by inspecting subpopulation differences among the HS&B students after they were put on the same scale as the NELS:88 cohort. If the linking worked as desired, then the relative differences that were found among the HS&B subpopulations on their original scales should not change when they were put on the new scaling. All subpopulation differences remained relatively invariant, indicating that the linking was successful.

In 1994, the IRT scales for all three waves of the survey were recalculated using different procedures. However, the NELS:88-HS&B mathematics test equating scales were not recalculated. Thus, the NELS:88-HS&B equated math scores are on the same scale as the original NELS:88 scores that were released with the first follow-up data tapes. While they are not comparable to the rescaled scores calculated in 1994, the Pearson correlation coefficients for the original versus the rescaled math test scores are greater than 0.99.

The NELS:88-HS&B equated math test scores for the 1980 HS&B sophomore cohort are available as a separate file for public use.

B.10. The NELS:88 1990 Census Data Files

Two kinds of zip code linkage to the 1990 Census have been effected in NELS:88. One linkage is at the school level. The second linkage is at the level of students' residential addresses.

B.10.1 School Zip Code Linkages

The school-level NELS:88 1990 Census data files contain selected 1990 zip code-level Census characteristics for the schools participating in the NELS:88 base-year, first follow-up and second follow-up school surveys. Census data aggregated at the zip code level (from the STF3B zip code-level Census files) were linked to NELS:88 schools by school zip code, which does not appear on any NELS:88 files. The NELS:88 Census variables are structural characteristics that are intended to approximate the local community surrounding the school. (No empirical mapping of school community boundaries compared to zip codes was undertaken for NELS:88). In the interest of standardization across zip codes, the raw counts provided in Census tables have, for many variables, been used to calculate the proportion of zip code residents displaying a given attribute (for example, the proportion of zip code residents who are Black). Researchers who wish to recalculate raw counts can easily do so using the data provided on the file.

The following variables characterizing the school's zip code are included on the files:

- number of housing units;
- number of residents;
- four separate variables providing the percentage of zip code residents living in areas classified as: 1) rural farm; 2) rural, not farm; 3) urban, in an urbanized area; or 4) urban, not in an urbanized area; it is not unusual for a single zip code to include residents with different urbanicity classifications;

- several ethnicity variables indicating the percentage of zip code residents who are White, Black, American Indian/Eskimo/Aleut, Asian or Pacific Islander, Hispanic (broken down into Mexican, Puerto Rican, Cuban and other Hispanic) or other ethnicity;
- variables indicating the proportion of zip code residents above and below the poverty level, by 12 age categories, as well as variables indicating the proportion of zip code residents with income-to-poverty ratios within defined ranges; and
- median income for the zip code.

NOTE: Researchers should note that, instead of attempting to characterize each school's zip code as urban **or** suburban **or** rural, as do the NELS:88 urbanicity variables, the Census scheme recognizes that diversity occurs even within small areas. It is not unusual to find that a single zip code encompasses residents with different urbanicity classifications; for example, one zip code may include some residents classified as rural-not farm and others classified as urban-not in an urbanized area.

See Ingels, Scott, and Taylor, *NELS:88 Base Year Through Second Follow-Up Final Methodology Report*, NCES Working Paper 98-06 (downloadable from the NCES Web Site) for a comprehensive listing of the 1990 Census measures available in NELS:88.

B.10.2 Residential Zip Code Linkages

Three special *student-level residential zip code Census variable files* have been created (1988, 1990, and 1992), and are available to licensed users on approval of special application. The data files contain 715 variables from 1990 Census Summary Tape File 3B (STF3B) linked to home zip codes for members of the 8th-grade cohort in 1988, 1990, and 1992. There are a variety of computed measures on population characteristics, labor force participation, education, fertility and marriage, and income/poverty. A few examples of some of the specific variables taken from the 1990 Census at the residence zip code level include: percent of families in poverty, median family income, percent of 25+ year olds graduated from college, percent of males unemployed (overall and by sex and race), percent of mothers with children in the labor force, ratio of single males to single females, percent of births to women under age 20, and so on. Primarily because zip code boundaries may change over time, there are a few schools (55 out of 2,487) and students (1,619 out of 64,000 records) that could not be matched to the Census variables.

As an alternative to the three files containing Census variables for the 1988-92 samples, a separate restricted-use file that links student ID to residential zip code can be requested. This file can be used by researchers to make their own selection of Census measures and to use their own programs to draw data from the 1990 and 2000 decennial Census data bases. Links of residential zip code to 1994 and 2000 NELS:88 data are of particular interest for exploring labor market issues. The fact that the fourth follow-up of NELS:88, occurring in the spring of the year that most sample members would turn 26, coincided with the 2000 Census, greatly enhances the utility of NELS:88 fourth follow-up data.

There are currently no linkages of NELS:88 data to decennial Census files at the tract level.

For linkage to the 2000 Decennial Census and other geocoded sources, the NELS:88/2000 restricted-use ECB includes the residential zip code of all fourth follow-up respondents.

B.11 NELS:88 QED-CCD-SDDB 1988-92 School Link Variables

The NELS:88 QED-CCD-SDDB school link files contain link variables that permit licensed researchers to merge the three waves of NELS:88 core school data with additional contextual variables on

the school and district frames available from Quality Education Data (QED), Inc., and NCES (the Common Core of Data [CCD] and the School District Data Book [SDDB]). The QED frames include records for public and private schools and public districts and Catholic dioceses. The CCD frame includes records for public schools and districts, while the SDDB files are at the public district (agency) level.

A wide range of information is available on the QED and CCD files. The QED files include information on grade span and enrollment size, the number of schools in a public district, instructional dollars per pupil, ethnic composition, urbanicity, and Orshansky percentile. FIPS county and metropolitan statistical area (MSA) codes are also provided. Variables that appear on CCD school and district files include: number of teachers per school, school enrollment, school racial/ethnic distribution, diplomas awarded, selected 1990 Census variables from the SDDB (available at the district level only) and financial information for districts extracted from the Survey of School District Finances data files.

The School District Data Book (SDDB), a CD-ROM product, is an unprecedented NCES resource for education research that provides thousands of 1990 Census variables and other data for all 15,274 public school districts in the United States. In collaboration with the Council of Chief State School Officers and the States, NCES contracted with the Census Bureau to map the geography of public school districts to the Census TIGER files. The 1990 Census variables were then retabulated within those geographic boundaries. Results are available at school district, county (FIPS state and county codes are provided), state and national levels. The SDDB also includes CCD data for the academic year 1989-1990 and data from the 1989-1990 Survey of School District Finances. The SDDB CD-ROM includes software for manipulating the data.

See appendix D of Ingels, Scott, and Taylor (NCES Working Paper 98-06) for detailed information on the NELS:88 QED-CCD-SDDB link variables.

B.12 NELS:88 QED District and School Data Files

A total of six district and school files—one school and one district file per wave—derived from files purchased from Quality Education Data (QED) of Denver, Colorado are available on the NELS:88 base-year to second follow-up restricted-use ECB. These files contain variables describing the characteristics of the public districts, Catholic dioceses and schools of all types that participated in the NELS:88 base-year, first follow-up and second follow-up surveys. The QED files include information on grade span and enrollment size, the number of schools in a public district and instructional dollars per pupil. (QED collects and sells a broad range of information on all schools in the United States, including private schools. In addition to the research community, the QED client base includes purveyors of educational goods such as textbook publishers and hardware/software vendors.) The QED data may be merged with the 1996 NELS:88 BY-F2 restricted-use school file, and subsequently the student-level file, for further investigation of contextual effects in the NELS:88 sample. The QED files may be merged with previously-released NELS:88 files using the NELS:88 QED-CCD-SDDB link file as a crosswalk.

The QED files have played an important role in NELS:88. The NELS:88 base-year district/diocesan and school sampling frames for institutions with 8th grades were compiled by QED. The files used in the NELS:88 base year were leased from QED in 1987. In 1989, QED files were leased for the first follow-up, and in 1991 for the second follow-up. In the first and second follow-ups, the QED files were used not for sampling but were used as sources of contacting and locating information for districts and schools to which sampled NELS:88 students had dispersed by 1990 and 1992. QED itself maintains only files with current information; the files used in NELS:88 are no longer available from QED. QED has generously given NCES and NORC permission to release the QED data for NELS:88

schools and their districts/dioceses to researchers. Detailed documentation on the NELS:88 QED district and school files is included as appendix E of Ingels, Scott, and Taylor, NCES 98-06.

B.13 NELS:88 Postsecondary Education Transcript Files

The Postsecondary Education Transcript Study (PETS) was designed to obtain official records from academic and vocational schools. Transcript information was abstracted and coded into machine-readable form, and can be merged with questionnaire data to support powerful quantitative analyses of the impacts of postsecondary schooling, or the impact of secondary school coursetaking, achievement and experience on postsecondary participation, coursetaking patterns and outcomes. Data files to be created for the transcript study include detailed information about program enrollments, periods of study, fields of study pursued, specific courses taken, grades received, and credentials earned. NELS:88 PETS files are restricted use only and require a license from NCES.

B.14 Other NELS:88/94 and NELS:88/2000 Special Restricted Variables

Certain information collected in the third and fourth follow-ups such as, in particular, many of the variables pertaining to postsecondary institutions, and some of the verbatims collected in telephone interviews, are available only on the restricted-use ECBs for 1988-1994 and 1988-2000 (N4R and N0R). In addition, the NELS:88/2000 restricted-use ECB (N0R) includes derived variables that summarize student postsecondary education financial aid information, including Stafford and Pell grant data.

Appendix C:
Response Rates, 1988-2000

Table C-1.—NELS:88 base-year school sample selection and realization: 1988

Stratum	Estimated[1] size	Eligible original selections	Target N	Total N cooperating schools	Sample realization (% of target achieved)	Cooperating original selections	Cooperative alternative selections
Total	38,837	1,002	1,032	1,057[2]	102%	698	359
Public schools[3]	22,690	774	800	817	102%	522	295
Catholic schools[3]	6,928	91	95	104	109%	70	34
Other private schools	9,219	137	137	136	99%	106	30

[1]Estimated as the sum of the school-level weights for each school type.

[2]1,057 schools participated at some level, though usable student data were received for only 1,052.

[3]Stratified by nine Census divisions; racial compositions; grade 8 enrollment; and urbanicity (central city, suburban within SMSA county, rural [non SMSA]).

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

Table C-2.—NELS:88 base-year sample member status and completion rates for student, parent, teacher, and school surveys, *adjusted for out-of-scope sample members*: 1988

	Student questionnaire completion rates		Student 8th grade test[1] completion rates		Parent questionnaire[2] completion rates		Teacher ratings[3] coverage rates		School questionnaire[4] completion rates	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Sample member status										
Participated		24,599		23,701		22,651		23,188		1,035
Selected		26,432		24,599		24,599		24,599		1,052
Completion rates										
Total	93.41	93.05	96.53	96.35	93.70	92.08	95.91	94.26	98.92	98.38
School type										
Public	93.15	92.79	96.32	96.11	94.21	93.72	96.57	95.82	98.73	98.28
Catholic	95.67	94.99	98.08	97.52	89.85	83.55	90.95	84.76	100.00	100.00
Other private	94.06	93.15	97.34	96.94	91.57	88.34	93.18	92.11	98.25	97.74
Urbanicity										
Urban	92.36	92.19	95.89	95.96	91.48	90.00	94.62	93.20	98.94	97.48
Suburban	92.17	92.38	96.36	96.29	93.32	91.44	95.56	93.85	98.12	98.18
Rural	95.26	95.13	97.29	96.94	96.08	95.40	97.46	96.09	99.64	99.66
Region										
Northeast	92.81	91.85	96.31	95.52	90.58	84.45	91.75	86.42	98.67	97.72
South	94.11	94.03	96.93	96.92	95.93	95.87	97.44	97.00	99.19	98.89
North central	94.70	94.79	96.85	96.96	94.92	94.72	97.71	97.82	99.75	98.88
West	91.17	90.83	95.50	95.40	90.18	89.62	94.18	93.25	97.10	97.54
Ethnicity										
Hispanic	90.86	90.24	94.95	94.88	88.35	87.57	92.58	92.50	†	†
Asian/Pacific Islander	89.70	90.12	98.18	97.84	90.76	91.53	94.06	93.69	†	†
Other	93.75	93.63	96.64	96.45	94.28	92.72	96.28	94.53	†	†
Minority schools										
>19 % minority students	89.64	89.43	95.21	95.44	89.94	88.79	92.78	92.44	98.54	98.04
≤19 % minority students	93.83	93.51	96.67	96.45	94.09	92.47	96.24	94.48	98.93	98.42

† Not applicable.

[1] 8th-grade cognitive test coverage rate for each student who completed a student questionnaire. In other words, 96.5% of students who completed the base-year student questionnaire also completed the cognitive tests.

[2] 8th-grade parent questionnaire coverage rate for each student who completed a student questionnaire.

[3] 8th-grade teacher rating coverage rate is based on the percentage of participating students for whom observations were obtained from one or more teachers.

[4] 8th-grade school completion rate (for school questionnaire), where at least one student completed a student questionnaire.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

Table C-3.—NELS:88 base-year sample member status and completion rates for student, parent, teacher, and school surveys, *with completions as a proportion of the total initial sample*: 1988

	Student questionnaire completion rates		Student 8th grade test [1] completion rates		Parent questionnaire[2] completion rates		Teacher ratings [3] completion rates		School questionnaire[4] completion rates	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Sample member status										
Participated		24,599		23,701		22,651		23,188		1,035
Selected		26,432		26,432		26,432		26,432		1,052
Completion rates										
Total	93.41	93.05	90.17	89.65	87.53	85.68	89.59	87.72	98.92	98.38
School type										
Public	93.15	92.79	89.73	89.18	87.75	86.97	89.95	88.92	98.73	98.28
Catholic	95.67	94.99	93.83	92.63	85.96	79.37	87.01	80.51	100.00	100.00
Other private	94.06	93.15	91.56	90.29	86.14	82.27	87.65	85.79	98.25	97.74
Urbanicity										
Urban	92.36	92.19	88.56	88.46	84.49	82.97	87.39	85.92	98.94	97.48
Suburban	92.17	92.38	89.34	88.96	86.52	84.47	88.60	86.70	98.12	98.18
Rural	95.26	95.13	92.68	92.14	91.52	90.74	92.85	91.41	99.64	99.66
Region										
Northeast	92.81	91.85	89.39	87.73	84.06	77.56	85.15	79.37	98.67	97.72
South	94.11	94.03	91.23	91.14	90.28	90.14	91.71	91.21	99.19	98.89
North central	94.70	94.79	91.71	91.91	89.89	89.78	92.53	92.72	99.75	98.88
West	91.17	90.83	87.07	86.69	82.21	81.40	85.87	84.69	97.10	97.54
Ethnicity										
Hispanic	90.86	90.24	86.27	85.63	80.28	79.02	84.11	83.48	†	†
Asian/Pacific Islander	89.70	90.12	88.07	88.17	81.41	82.49	84.37	84.43	†	†
Other	93.75	93.63	90.61	90.31	88.39	86.81	90.26	88.51	†	†
Minority schools										
>19 % minority students	89.64	89.43	85.35	85.36	80.63	79.41	83.17	82.67	98.54	98.04
≤19 % minority students	93.83	93.51	90.70	90.19	88.29	86.47	90.30	88.35	98.93	98.42

† Not applicable.

[1] 8th-grade cognitive test coverage rate for this table is based on test completion rates, regardless of student completion of the student questionnaire.

[2] 8th-grade parent questionnaire coverage rate is based on questionnaire completion by the parents, regardless of student participation.

[3] 8th-grade teacher rating coverage rate is based on the percentage of all sampled students for whom observations were obtained from one or more teachers.

[4] 8th grade completion rate for the school questionnaire, regardless of student participation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988.

Table C-4.—NELS:88 base-year completion rates by sample eligibility for *base-year sample members retained in the first follow-up*: 1988-1990

	Student questionnaire completion rates		Student 8th grade test[1] completion rates		Parent questionnaire[2] completion rates		Teacher ratings[3] completion rates		School questionnaire[4] completion rates	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Total	93.95	93.63	96.54	96.32	94.69	93.47	96.33	95.09	98.67	98.14
School type										
Public	93.81	93.52	96.42	96.15	95.06	94.69	96.96	96.40	98.52	98.03
Catholic	95.68	94.65	97.75	97.21	91.13	86.04	89.78	85.25	100.00	100.00
Other private	94.89	93.78	97.52	97.09	90.71	88.80	90.24	91.54	97.14	97.37
Urbanicity										
Urban	92.86	92.82	95.62	95.76	92.40	91.26	95.24	94.32	98.57	97.08
Suburban	93.09	92.71	96.52	96.41	94.55	93.13	96.00	94.84	97.82	97.91
Rural	95.73	95.61	97.08	96.66	96.20	95.80	97.38	96.07	99.57	99.65
Region										
Northeast	93.81	92.59	96.12	95.28	92.45	87.07	93.35	88.73	98.57	97.66
South	93.76	94.00	96.56	96.58	95.71	95.46	98.46	98.53	98.74	98.31
North central	95.50	95.37	97.39	97.23	96.74	96.79	96.83	95.98	99.71	98.83
West	92.27	91.77	95.68	95.66	92.07	91.71	94.57	93.94	96.54	97.44
Ethnicity										
Hispanic	92.60	91.77	95.07	95.11	90.10	89.05	92.38	92.01	†	†
Asian/Pacific Islander	92.67	91.95	96.38	96.94	90.30	91.25	95.44	94.49	†	†
Black	94.29	94.72	95.12	95.06	92.15	91.75	96.19	95.53	†	†
White	95.81	95.68	96.91	96.64	96.25	95.14	96.96	95.72	†	†
American Indian	87.97	87.45	99.07	98.61	78.25	75.00	93.66	91.20	†	†
Minority schools[5]										
>19 % minority students	91.61	91.41	95.56	95.89	90.96	90.49	93.90	93.44	98.54	98.04
≤19 % minority students	94.17	93.87	96.63	96.37	95.04	93.79	96.55	95.27	98.67	98.15

† Not applicable.

[1] 8th-grade cognitive test coverage rate for each student who completed a student questionnaire.

[2] 8th-grade parent questionnaire coverage rate for each student who completed a student questionnaire.

[3] 8th-grade teacher rating completion rate is based on the percentage of participating students for whom observations were obtained from one or more teachers

[4] 8th-grade school questionnaire coverage rate for each student who completed a student questionnaire.

[5]Refers to 8th-grade schools.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1990.

Table C-5.—NELS:88 first follow-up completion rates (10th grade cross-section) by sample eligibility: 1990

	Student questionnaire completion rates		Student 10th grade test[1] completion rates		Dropout questionnaire completion rates		Dropout 10th grade test[2] completion rates		School questionnaire[3] completion rates		School questionnaire[4] completion rates	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Total	91.09	94.10	94.14	95.23	90.97	89.84	48.56	50.05	†	97.07	91.97	96.94
School type[e]												
Public	91.66	94.38	94.34	95.39	†	†	†	†	†	97.41	93.20	97.28
Catholic	97.53	97.62	95.22	97.05	†	†	†	†	†	95.90	88.95	95.22
Other private	89.51	93.27	91.64	93.53	†	†	†	†	†	95.16	82.77	97.89
Urbanicity[5]												
Urban	90.36	93.64	92.29	93.53	†	†	†	†	†	96.65	90.95	96.90
Suburban	92.25	94.53	94.80	95.91	†	†	†	†	†	96.94	92.97	97.19
Rural	93.31	95.73	95.91	96.66	†	†	†	†	†	98.76	94.17	98.11
Region[5]												
Northeast	91.84	93.26	93.57	94.32	†	†	†	†	†	95.10	93.83	96.87
South	93.09	95.78	94.68	96.12	†	†	†	†	†	97.82	91.43	97.18
North central	93.60	95.42	97.22	97.45	†	†	†	†	†	98.46	94.70	98.58
West	87.46	92.02	90.02	92.08	†	†	†	†	†	96.17	90.17	95.80
Ethnicity												
Asian/Pacific Islander	90.71	92.96	93.59	94.64	70.37	75.00	23.77	28.57	†	†	94.63	97.28
Hispanic	88.32	92.75	90.18	92.54	91.72	87.64	43.81	50.22	†	†	89.46	94.39
Black	88.85	93.89	92.13	94.02	89.02	87.10	39.41	48.77	†	†	87.92	95.88
White	93.56	95.69	95.14	96.02	93.78	94.06	55.26	52.39	†	†	92.95	97.55
American Indian	88.46	92.15	97.78	97.76	88.62	83.33	40.46	36.00	†	†	93.65	97.31
Refused/missing	28.92	35.52	80.40	80.43	66.25	62.86	27.72	31.82	†	†	†	†

† Not applicable.
[1] 10th grade cognitive test coverage rate for each student who completed a student questionnaire.
[2] 10th grade cognitive test coverage rate for each dropout who completed a dropout questionnaire.
[3] 10th grade school completion rate (for school questionnaire), where at least one student completed a student questionnaire.
[4] 10th grade school questionnaire coverage rate for each student who completed a student questionnaire.
[5] Refers to 10th-grade schools.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1990.

Table C-6.—NELS:88 *combined* base-year and first follow-up completion rates (panel members) by sample eligibility for student/dropout and parent surveys: 1988-1990

	Student/dropout questionnaire (both BY and F1) completion rates		Student/dropout cognitive test[1] (both BY and F1) Completion rates		Student/dropout cognitive test[1] (BY and/or F1) completion rates		Parent questionnaire[2] (BY only) completion rates	
	Weighted	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**	**Unweighted**
Total	92.77	95.42	89.05	90.47	99.53	99.66	94.32	94.00
School type[3]								
Public	92.43	95.37	88.50	90.00	99.54	99.67	94.77	95.17
Catholic	95.24	96.12	93.82	93.72	99.23	99.63	90.44	86.61
Other private	94.84	95.25	91.11	91.91	99.85	99.64	92.61	89.67
Urbanicity[3]								
Urban	91.02	94.39	84.89	88.32	99.02	99.60	92.31	92.05
Suburban	92.29	94.85	89.61	90.65	99.65	99.63	94.44	93.69
Rural	94.94	97.05	91.67	91.98	99.78	99.75	95.80	96.00
Region[3]								
Northeast	93.09	94.51	88.90	89.55	99.63	99.60	91.77	87.90
South	93.86	96.61	87.97	90.46	99.25	99.61	95.66	95.10
North central	94.35	96.18	93.85	94.07	99.74	99.78	96.73	97.18
West	88.28	93.16	84.34	86.45	99.67	99.64	90.95	92.45
Ethnicity								
Asian/Pacific Islander	90.68	93.87	87.65	90.53	99.99	99.91	91.32	91.86
Hispanic	89.38	93.73	84.83	86.38	99.56	99.58	89.96	89.87
Black	88.48	93.44	81.59	86.98	98.62	99.55	90.90	92.47
White	94.30	96.23	91.03	91.71	99.68	99.68	96.08	95.51
American Indian	87.36	91.16	91.36	90.31	99.38	99.49	76.80	76.53
Refused/missing	83.98	92.86	53.41	69.23	93.10	92.31	00.00	00.00
Minority schools[3]								
> 19 % minority students	85.87	92.69	79.63	83.14	99.72	99.76	90.98	91.45
≤ 19 % minority students	93.54	95.71	90.02	91.23	99.51	99.65	94.67	94.26

[1]Cognitive test coverage for each sample member who completed a BY student questionnaire and F1 student/dropout questionnaire.

[2]BY parent questionnaire coverage rate for each sample member who completed a BY student questionnaire and 1F student/dropout questionnaire.

[3]Refers to 8th-grade schools.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1990.

Table C-7.—NELS:88 *combined* base-year and first follow-up completion rates (panel members) by sample eligibility for the student (only) and school surveys: 1988-1990

	Student questionnaire (both BY and 1F) completion rates		School questionnaire[1] (both BY and F1) completion rates		School questionnaire[1] (BY and/or F1) completion rates	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Total	92.57	95.41	90.59	95.68	99.88	99.91
School type[2]						
Public	92.19	95.36	91.45	95.58	99.86	99.89
Catholic	95.19	96.07	87.77	95.75	100.00	100.00
Other private	94.83	95.24	81.11	96.40	100.00	100.00
Urbanicity[2]						
Urban	90.68	94.37	85.08	93.50	99.83	99.74
Suburban	92.10	94.86	90.25	95.03	99.82	99.94
Rural	94.83	97.02	95.51	98.32	100.00	100.00
Region[2]						
Northeast	92.88	94.44	91.52	95.57	99.96	99.97
South	93.58	96.57	90.36	95.98	99.85	99.97
North central	94.34	96.18	92.47	97.84	99.77	99.75
West	88.01	93.31	87.26	92.28	99.99	99.97
Ethnicity						
Asian/Pacific Islander	90.74	94.03	90.06	93.85	99.90	99.90
Hispanic	88.77	93.65	85.89	91.30	99.64	99.80
Black	87.92	93.56	86.03	94.56	99.94	99.94
White	94.16	96.17	91.99	96.73	99.89	99.92
American Indian	86.69	91.33	91.58	95.53	100.00	100.00
Refused/missing	78.10	91.67	100.00	100.00	100.00	100.00
Minority schools[2]						
> 19 % minority students	85.13	92.89	85.35	89.52	†	100.00
≤ 19% minority students	93.39	95.67	91.12	96.31	†	99.00

† Not applicable.

[1]School questionnaire coverage rate for each student who has completed a BY student questionnaire and F1 student questionnaire.

[2]Refers to 8th grade schools.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1990.

Table C-8.—NELS:88 second follow-up component survey completion rates by selected characteristics: 1992

	Student sample completion rates		Student 12th grade test[1] completion rates		Dropout/alternative[2] sample completion rates		Dropout /alt. 12th grade test[3] completion rates		School questionnaire[4] completion rates		School questionnaire[5] completion rates	
	Wgt	Unwgt	Wgt	Unwgt	Wgt	Unwgt	Wgt	Unwgt	Wgt	Unwgt	Wgt	Unwgt
Total	91.0	92.5	76.6	78.8	88.0	87.6	41.7	40.3	†	97.1	98.3	98.2
School type[6]												
Public	94.7	95.3	76.8	78.9	†	†	†	†	†	97.2	98.4	98.4
Catholic	98.4	98.0	79.7	84.5	†	†	†	†	†	97.1	96.6	96.7
Other private	94.8	95.5	73.1	75.6	†	†	†	†	†	96.0	98.5	97.2
Urbanicity[6]												
Urban	95.0	95.8	73.6	76.7	†	†	†	†	†	97.0	98.2	98.3
Suburban	94.4	95.2	74.9	75.7	†	†	†	†	†	97.4	98.5	98.2
Rural	95.5	95.5	82.4	85.3	†	†	†	†	†	96.6	99.8	98.0
Region[6]												
Northeast	94.3	94.7	77.6	76.7	†	†	†	†	†	94.7	97.9	96.8
South	95.4	95.8	77.7	81.7	†	†	†	†	†	97.3	98.2	98.4
Midwest	96.1	95.8	78.6	80.7	†	†	†	†	†	97.8	98.5	98.7
West	92.9	95.4	72.2	74.2	†	†	†	†	†	98.3	98.7	98.6
Ethnicity												
Asian/Pacific Islander	91.7	92.7	75.2	75.5	74.7	82.4	47.6	35.7	†	†	98.2	98.9
Hispanic	86.6	89.8	73.9	76.6	88.3	87.5	35.6	36.1	†	†	98.8	98.9
Black	88.1	90.5	74.6	77.1	84.8	83.6	37.2	38.7	†	†	98.3	98.0
White	93.5	94.2	77.8	80.1	89.7	89.5	44.2	42.4	†	†	98.3	98.0
American Indian	90.3	86.5	74.0	74.3	97.6	95.8	51.5	49.3	†	†	98.7	98.7
Refused/missing[7]	28.5	33.2	22.2	31.1	55.9	61.5	23.5	25.0	†	†	97.9	97.8

† Not applicable—completion rates by school type, urbanicity, and region are calculated based on the school a student attended in the second follow-up. Because dropouts are not linked to schools on the public-use magnetic tape, it is not possible to calculate dropout completion rates for these subgroups.

[1] 12th grade cognitive test coverage rate for each student who completed a questionnaire.

[2] Alternative completers could have completed either a student or dropout questionnaire, depending on status during data collection; 350 alternative sample members completed a student questionnaire, and 457 completed a dropout questionnaire.

[3] 12th grade cognitive test coverage rate for each dropout who completed a student questionnaire.

[4] 12th grade school completion rate (for school questionnaire) of eligible contextual schools, where at least one student completed a questionnaire.

[5] 12th grade school questionnaire coverage rate for each student who completed a questionnaire and was enrolled in an eligible contextual school.

[6] Refers to second follow-up school.

[7] Refused/missing refers only to the status of a sample member's ethnicity. It does not refer to sample members who did not participate in the second follow-up.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1992.

Table C-9.—NELS:88 second follow-up completion rates for base-year to first follow-up panel participants by selected characteristics:[1] 1988-1990

	BY, F1, and F2 Student/dropout questionnaire completion rates		BY, F1, and F2 student/dropout cognitive test completion rates[2]	
	Weighted	**Unweighted**	**Weighted**	**Unweighted**
Total	94.7	95.1	69.6	72.2
School type[3]				
Public	94.3	94.7	69.0	71.4
Catholic	97.9	97.0	74.1	78.6
Other private	97.4	97.0	73.0	73.7
Urbanicity[3]				
Urban	93.5	95.1	64.3	69.5
Suburban	95.5	95.3	69.1	70.1
Rural	94.8	94.9	74.6	77.2
Region[3]				
Northeast	94.8	95.1	70.3	71.3
South	94.1	94.5	68.2	73.1
Midwest	95.7	96.0	74.9	76.4
West	94.6	95.1	63.7	65.7
Ethnicity				
Asian/Pacific Islander	93.3	95.0	71.5	71.9
Hispanic	93.1	94.4	63.9	65.5
Black	92.4	92.6	59.6	67.0
White	95.5	95.7	72.1	74.2
American Indian	94.1	91.3	64.8	64.0
Refused/missing[4]	81.1	75.0	38.3	55.6
Minority schools[3]				
> 19 % minority students	92.2	93.5	55.1	59.3
≤ 19 % minority students	95.0	95.3	71.0	73.5

[1] These panel completion rates are the proportion of base-year-first follow-up completers for whom a second follow-up questionnaire was completed but excludes base-year nonparticipants.

[2] Cognitive test coverage rate for each sample member who completed a BY student questionnaire, F1 and F2 student/dropout questionnaire.

[3] Refers to 8th-grade schools.

[4] Refused/missing refers only to the status of a sample member's ethnicity. It does not refer to sample member nonparticipants.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1990.

Table C-10.—NELS:88 second follow-up completion rates for base-year to first follow-up panel participants, by selected characteristics:[1] 1988-1992

	BY, F1, and F2 student questionnaire completion rates		BY, F1, and F2 school questionnaire completion rates[2]	
	Weighted	**Unweighted**	**Weighted**	**Unweighted**
Total	95.7	96.1	95.5	95.6
School type[3]				
Public	95.4	95.8	95.8	95.7
Catholic	98.2	97.3	94.3	94.8
Other private	97.5	97.1	93.5	95.8
Urbanicity[3]				
Urban	94.4	96.4	93.7	94.7
Suburban	96.2	96.1	94.4	94.3
Rural	95.8	95.9	98.4	98.2
Region[3]				
Northeast	95.2	95.5	94.9	94.6
South	95.8	96.2	95.6	95.9
Midwest	96.2	96.5	97.5	97.8
West	95.5	96.0	93.1	93.2
Ethnicity				
Asian/Pacific Islander	94.9	95.8	90.2	93.9
Hispanic	94.2	95.8	89.8	91.3
Black	94.3	95.0	95.1	95.3
White	96.2	96.4	96.5	96.5
American Indian	93.8	90.9	97.6	97.3
Refused/missing[4]	74.2	72.7	100.0	100.0
Minority schools[3]				
> 19 % minority students	92.5	96.3	90.7	90.0
≤ 19 % minority students	96.0	94.4	96.0	96.2

[1] These panel completion rates are the proportion of base-year-first follow-up completers for whom a second follow-up questionnaire was completed but excludes base-year nonparticipants.

[2] School questionnaire coverage rate for each student who completed a BY student questionnaire, F1 and F2 student questionnaire.

[3] Refers to 8th-grade schools.

[4] Refused/missing refers only to the status of a sample member's ethnicity. It does not refer to sample member nonparticipants.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1988-1992.

Table C-11.—NELS:88 third follow-up study completion rates, by selected sampling strata: 1994

	Total	Completed cases	Unweighted percent	Weighted percent
Total	15,875[1]	14,915	93.95	90.86
Respondent sex				
Male	7,895	7,354	93.15	90.24
Female	7,980	7,561	94.75	91.48
Respondent race/ethnicity				
Asian/Pacific Islander	1,151	1,088	94.53	90.85
Hispanic	2,288	2,107	92.09	87.98
Black	1,840	1,681	91.36	87.45
White	10,303	9,787	94.99	92.90
American Indian	230	211	91.74	91.86
Missing	63	41	65.08	47.87
Second follow-up standard test quartile				
Lowest test quartile	2,669	2,497	93.56	91.98
2nd	2,850	2,710	95.09	94.21
3rd	2,836	2,746	96.83	96.71
4th	2,982	2,923	98.02	98.09
Missing	55	53	96.36	98.54
Did not complete test	4,483	3,986	88.91	82.62
Socioeconomic status quartile				
Lowest SES quartile	4,062	3,788	93.25	91.78
2nd	3,784	3,587	94.79	93.56
3rd	3,742	3,570	95.40	94.75
4th	3,635	3,507	96.48	96.42
Missing	652	463	71.01	64.61
Sample type				
8th-grade cohort	14,890	14,041	94.30	91.48
F2 freshened	117	102	87.18	76.88
F1 freshened	559	501	89.62	85.14
Base year ineligible	309	271	87.70	81.50
Dropout status				
Never dropped out	13,337	12,654	94.88	92.38
Ever dropped out	2,538	2,261	89.09	83.77
Original school type				
Public	13,383	12,540	93.70	90.59
Catholic	1,355	1,292	95.35	93.89
NAIS private	595	568	95.46	89.37
Other private	542	515	95.02	92.12

Table C-11.—NELS:88 third follow-up study completion rates, by selected sampling strata: 1994—Continued

	Total	Completed cases	Unweighted percent	Weighted percent
Sample strata[2]				
Nonresponders	38	9	23.68	24.97
Poor responders	595	444	74.62	75.81
Dropouts	2,343	2,133	91.04	89.36
Ineligible prior to 1992	191	176	92.15	91.80
Private school in 1988	2,370	2,269	95.74	95.66
Private school in 1990/1992	96	92	95.83	98.12
Hispanic	1,457	1,376	94.44	95.38
Asian/Pacific Islander	870	833	95.75	95.90
American Indian	132	125	94.70	94.09
Black high test quartile	79	75	94.94	96.56
Black other	1,112	1,034	92.99	90.20
White low SES quartile	1,292	1,228	95.05	93.87
White high SES quartile	1,505	1,472	97.81	97.63
White mid SES	3,789	3,644	96.17	95.61
F1 freshened	1	1	100	100
F2 freshened	2	2	100	100
Other	3	2	66.67	66.67

[1] This does not include 89 ineligible or deceased sample members.

[2] The third follow-up sampling strata were not mutually exclusive, and sample members were assigned to strata in priority order. Thus, "other" respondents were not members of any of the previous sample strata.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 1994.

Table C-12.—NELS:88 fourth follow-up study completion rates, by selected sample strata: 2000

Strata	Total	Completed cases	Unweighted percent	Weighted percent
Total	15,649[1]	12,144	77.60	82.71
Respondent sex				
Male	7,715	5,782	74.94	79.55
Female	7,934	6,362	80.19	86.06
Respondent race/ethnicity				
Asian/Pacific Islander	1,120	853	76.16	79.71
Hispanic	2,253	1,622	71.99	74.51
Black	1,788	1,176	65.77	83.06
White	10,191	8,322	81.66	84.31
American Indian	213	161	75.59	76.35
Missing	84	10	11.90	91.38
Second follow-up standard test quartile				
Lowest test quartile	2,626	1,860	70.83	83.83
2nd	2,822	2,208	78.24	84.06
3rd	2,857	2,404	84.14	89.01
4th	2,965	2,687	90.62	92.54
Missing	10	9	90.00	100.00
Did not complete test	4,369	2,976	68.12	72.68
Socioeconomic status quartile				
Lowest SES quartile	3,889	2,797	71.92	76.38
2nd	3,541	2,737	77.29	84.28
3rd	3,475	2,794	80.40	85.12
4th	3,517	3,056	86.89	89.08
F1/F2 Freshened students	685	401	58.54	64.36
Missing	542	359	66.24	81.82
Sample type				
8th-grade cohort	14,672	11,559	78.78	83.75
F2 freshened	142	63	44.37	73.79
F1 freshened	543	338	62.25	63.17
Base-year ineligible	292	184	63.01	79.30
Dropout status				
Never dropped out	13,202	10,556	79.96	83.53
Ever dropped out	2,447	1,588	64.90	77.78
8th-grade school type				
Public	12,506	9,693	77.51	83.32
Catholic	1,348	1,114	82.64	82.21
Private, other religion	465	400	86.02	93.18
Private, non-religious	594	509	85.69	90.84
F1/F2 Freshened students	685	401	58.54	64.36
Missing	51	27	52.94	73.80

Table C-12.—NELS:88 fourth follow-up study completion rates, by selected sample strata: 2000—Continued

Strata	Total	Completed cases	Unweighted percent	Weighted percent
Third follow-up study sampling strata[2]				
Nonresponders	9	7	77.78	100.00
Poor responders	588	305	51.87	58.97
Dropouts	2,260	1,512	66.90	87.62
Ineligible prior to 1992	183	125	68.31	82.72
Private school in 1988	2,351	1,967	83.67	86.00
Private school in 1990/1992	96	84	87.50	91.28
Hispanic	1,445	1,079	74.67	84.77
Asian/Pacific Islander	862	664	77.03	81.52
American Indian	131	95	72.52	74.25
Black high test quartile	79	62	78.48	100.00
Black other	1,098	736	67.03	88.48
White low SES quartile	1,284	1,022	79.60	80.72
White high SES quartile	1,497	1,336	89.25	90.43
White mid SES	3,760	3,146	83.67	84.72
F1 freshened	1	0	0.00	0.00
F2 freshened	2	2	100.00	100.00
Other	3	2	66.67	100.00

[1]This does not include 315 out-of-scope or deceased sample members.

[2]The third follow-up sampling strata are not mutually exclusive, and sample members were assigned to strata in priority order. Thus, "other" respondents were not members of any of the previous sample strata.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Appendix D: Design Effect Tables for NELS:88/2000

Appendix D
Design Effect Tables for NELS:88/2000

Tables D–1 through D–10 in this appendix describe the precision of the estimates only for the NELS:88 fourth follow-up study. Similar design effect tables for the base-year and first three follow-up studies have been published and discussed elsewhere and are not repeated here. Interested readers can find these estimates in the following publications.

Base-year and first and second follow-up studies

See the discussion in section 3.3 of Ingels, S.J., Dowd, K.L., Baldridge, J.D., Stipe, J.L., Bartot, V.H., and Frankel, M.R. (1994). *NELS:88 Second Follow-Up Student Component Data File User's Manual* (NCES 94-374). Washington, DC: NCES, for the base-year and first two follow-up studies. In the same report, tables 3.3-1 and 3.3-2 report design effect estimates for the base-year study; tables 3.3-5 and 3.3-6 describe the average design effects for the first follow-up study; and tables 3.3.12 through 3.3.14 depict estimates for the second follow-up. Additional standard error and design effect tables for the base year through second follow-up may be found in appendix F of the *NELS:88 Base Year Through Second Follow-Up Final Methodology Report* (Ingels, S.J., Scott, L.A., and Taylor, J.R.; NCES 98-06) (this publication is available on the NCES Web Site).

Third follow-up study

Interested data users should review section 5.3 in Haggerty, C., Dugoni, B.L., Reed, L., Cederlund, A., and Taylor, J.R. (1996). *Methodology Report: National Education Longitudinal Study: 1988-1994* (NCES 96-174). Washington, DC: NCES, for a discussion of the average design effects for NELS:88/94.

Table D-1.—Standard errors and design effects for *all respondents* to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	16.312	10,636	0.789	4.855	2.203	0.358
How many times did R cut/skip classes	F2S9B	4.468	10,643	0.471	5.533	2.352	0.200
How far in school mother wants R to go	F2S42B	39.764	9,990	0.868	3.140	1.772	0.490
R feels good about him/herself	F2S66A	93.516	9,865	0.371	2.240	1.497	0.248
# R's friends to attend 4 yr school	F2S69E	54.571	9,893	1.028	4.220	2.054	0.501
Number of jobs during period-#1	NUMJOBS1	1.515	9,486	0.014	2.365	1.538	0.009
Time spent working on hobbies	HOBBIES	51.990	12,013	0.773	2.873	1.695	0.456
Respondent/family member seriously ill	ILLDISBL	27.979	11,996	0.730	3.168	1.780	0.410
Importance of success in work	SUCSLWRK	89.681	12,013	0.457	2.708	1.646	0.278
PSE: Number valid PSE institutions reported	F3PSENUM	0.771	12,051	0.012	4.044	2.011	0.006
Current activity-full-time job	F4AACTF	76.622	12,143	0.689	3.220	1.794	0.384
January 2000 activities same as current	F4AACTD	92.768	11,889	0.431	3.288	1.813	0.238
Ever served in armed forces	F4AAFTN	93.495	11,964	0.414	3.376	1.837	0.225
How got current job	F4BHGJ	38.396	11,824	0.823	3.382	1.839	0.447
Employed for at least 6 months in 1998	F4BJ98A	11.290	11,670	0.511	3.046	1.745	0.293
Perceived job autonomy	F4BJAUT	33.128	12,004	0.720	2.808	1.676	0.430
Received license since high school	F4CLICQ	71.626	11,942	0.841	4.152	2.038	0.413
Training received in last 12 months	F4CTRNQ	43.484	11,210	0.879	3.521	1.876	0.468
Attended any PSE since high school	F4EANY	54.969	12,121	0.892	3.895	1.974	0.452
Primarily student or employee	F4ESTEM	77.509	9,126	1.035	5.609	2.368	0.437
PSE impact-better jobs	F4EJOBS	25.353	9,393	0.895	3.979	1.995	0.449
Participate in adult education	F4FAEQRY	88.494	11,920	0.494	2.855	1.690	0.292
Marital status in 2000	F4GMRS	39.677	12,101	0.889	3.995	1.999	0.445
Total number of dependents	F4GNDP	0.765	12,089	0.024	6.008	2.451	0.010
Income of respondent in 1999	F4HI99	756.265	11,146	25.320	2.461	1.569	16.140
Housing status	F4HHOSE	30.222	11,888	0.806	3.664	1.914	0.421
Registered to vote	F4IRVOTE	77.960	11,859	0.722	3.599	1.897	0.381
Use of cigarettes	F4ISMOKE	26.895	11,890	0.787	3.741	1.934	0.407
Hispanic origin	F4JHISP	12.400	11,890	0.922	9.307	3.051	0.302
Diversity in present neighborhood	F4JRDVB	66.372	11,680	0.749	6.863	2.620	0.286
Minimum					2.240	1.497	
Maximum					9.307	3.051	
Mean					3.930	1.954	
Standard deviation					1.493	0.339	
Median					3.560	1.887	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-2.—Standard errors and design effects for *male respondents* to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	15.961	5074	1.028	3.994	1.998	0.514
How many times did R cut/skip classes	F2S9B	5.434	5078	0.826	6.742	2.597	0.318
How far in school mother wants R to go	F2S42B	40.609	4689	1.399	3.806	1.951	0.717
R feels good about him/herself	F2S66A	95.814	4661	0.387	1.744	1.320	0.293
# R's friends to attend 4 yr school	F2S69E	49.514	4605	1.593	4.676	2.162	0.737
Number of jobs during period-#1	NUMJOBS1	1.504	4682	0.020	2.295	1.515	0.013
Time spent working on hobbies	HOBBIES	58.521	5694	1.127	2.980	1.726	0.653
Respondent/family member seriously ill	ILLDISBL	26.736	5685	1.018	3.006	1.734	0.587
Importance of success in work	SUCSLWRK	89.436	5696	0.654	2.577	1.605	0.407
PSE: Number valid PSE institutions reported	F3PSENUM	0.728	5713	0.017	3.598	1.897	0.009
Current activity-full-time job	F4AACTF	84.552	5782	0.886	3.471	1.863	0.475
January 2000 activities same as current	F4AACTD	93.889	5676	0.507	2.539	1.594	0.318
Ever served in armed forces	F4AAFTN	88.761	5692	0.787	3.533	1.880	0.419
How got current job	F4BHGJ	41.806	5653	1.283	3.822	1.955	0.656
Employed for at least 6 months in 1998	F4BJ98A	7.224	5608	0.630	3.319	1.822	0.346
Perceived job autonomy	F4BJAUT	31.182	5743	1.100	3.239	1.800	0.611
Received license since high school	F4CLICQ	70.985	5685	1.371	5.187	2.277	0.602
Training received in last 12 months	F4CTRNQ	42.714	5500	1.394	4.366	2.090	0.667
Attended any PSE since high school	F4EANY	55.721	5771	1.433	4.800	2.191	0.654
Primarily student or employee	F4ESTEM	73.358	4216	1.841	7.307	2.703	0.681
PSE impact-better jobs	F4EJOBS	27.635	4348	1.410	4.320	2.078	0.678
Participate in adult education	F4FAEQRY	90.314	5676	0.702	3.194	1.787	0.393
Marital status in 2000	F4GMRS	34.988	5763	1.275	4.118	2.029	0.628
Total number of dependents	F4GNDP	0.636	5758	0.035	6.831	2.614	0.014
Income of respondent in 1999	F4HI99	625.301	5361	31.075	2.521	1.588	19.571
Housing status	F4HHOSE	27.421	5663	1.063	3.213	1.792	0.593
Registered to vote	F4IRVOTE	75.985	5644	1.165	4.200	2.049	0.569
Use of cigarettes	F4ISMOKE	29.650	5658	1.248	4.227	2.056	0.607
Hispanic origin	F4JHISP	12.239	5656	1.195	7.517	2.742	0.436
Diversity in present neighborhood	F4JRDVB	65.419	5553	1.184	7.970	2.823	0.419
Minimum					1.744	1.320	
Maximum					7.970	2.823	
Mean					4.170	2.008	
Standard deviation					1.619	0.379	
Median					3.814	1.953	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-3.—Standard errors and design effects for *female respondents* to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	16.672	5562	0.930	3.464	1.861	0.500
How many times did R cut/skip classes	F2S9B	3.474	5565	0.435	3.143	1.773	0.246
How far in school mother wants R to go	F2S42B	38.921	5301	1.034	2.383	1.544	0.670
R feels good about him/herself	F2S66A	91.200	5204	0.638	2.643	1.626	0.393
# R's friends to attend 4 yr school	F2S69E	59.569	5288	1.144	2.874	1.695	0.675
Number of jobs during period-#1	NUMJOBS1	1.527	4804	0.019	2.341	1.530	0.012
Time spent working on hobbies	HOBBIES	45.756	6319	1.004	2.567	1.602	0.627
Respondent/family member seriously ill	ILLDISBL	29.168	6311	1.015	3.145	1.773	0.572
Importance of success in work	SUCSLWRK	89.915	6317	0.632	2.778	1.667	0.379
PSE: Number valid PSE institutions reported	F3PSENUM	0.811	6338	0.016	3.639	1.908	0.009
Current activity-full-time job	F4AACTF	68.643	6361	0.971	2.783	1.668	0.582
January 2000 activities same as current	F4AACTD	91.630	6213	0.687	3.827	1.956	0.351
Ever served in armed forces	F4AAFTN	98.243	6272	0.199	1.437	1.199	0.166
How got current job	F4BHGJ	34.922	6171	0.969	2.549	1.596	0.607
Employed for at least 6 months in 1998	F4BJ98A	15.492	6062	0.818	3.097	1.760	0.465
Perceived job autonomy	F4BJAUT	35.118	6261	0.969	2.579	1.606	0.603
Received license since high school	F4CLICQ	72.270	6257	0.916	2.621	1.619	0.566
Training received in last 12 months	F4CTRNQ	44.318	5710	1.048	2.541	1.594	0.657
Attended any PSE since high school	F4EANY	54.213	6350	1.044	2.790	1.670	0.625
Primarily student or employee	F4ESTEM	81.405	4910	0.879	2.506	1.583	0.555
PSE impact-better jobs	F4EJOBS	23.204	5045	1.017	2.926	1.711	0.594
Participate in adult education	F4FAEQRY	86.666	6244	0.662	2.368	1.539	0.430
Marital status in 2000	F4GMRS	44.403	6338	1.063	2.903	1.704	0.624
Total number of dependents	F4GNDP	0.895	6331	0.026	3.670	1.916	0.014
Income of respondent in 1999	F4HI99	889.719	5785	40.351	2.522	1.588	25.409
Housing status	F4HHOSE	33.041	6225	1.038	3.031	1.741	0.596
Registered to vote	F4IRVOTE	79.943	6215	0.795	2.452	1.566	0.508
Use of cigarettes	F4ISMOKE	24.128	6232	0.974	3.230	1.797	0.542
Hispanic origin	F4JHISP	12.563	6234	0.984	5.498	2.345	0.420
Diversity in present neighborhood	F4JRDVB	67.325	6127	0.735	3.560	1.887	0.390
Minimum					1.437	1.199	
Maximum					5.498	2.345	
Mean					2.929	1.701	
Standard deviation					0.693	0.194	
Median					2.786	1.669	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-4.—Standard errors and design effects for *Asian and Pacific Islander respondents* to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	23.588	797	2.857	3.604	1.899	1.505
How many times did R cut/skip classes	F2S9B	4.978	805	1.284	2.803	1.674	0.767
How far in school mother wants R to go	F2S42B	34.699	755	3.015	3.026	1.739	1.733
R feels good about him/herself	F2S66A	90.244	751	1.784	2.711	1.647	1.083
# R's friends to attend 4 yr school	F2S69E	65.263	745	2.943	2.843	1.686	1.745
Number of jobs during period-#1	NUMJOBS1	1.375	571	0.047	2.187	1.479	0.032
Time spent working on hobbies	HOBBIES	54.879	841	3.195	3.462	1.861	1.717
Respondent/family member seriously ill	ILLDISBL	19.351	838	2.093	2.349	1.533	1.365
Importance of success in work	SUCSLWRK	86.381	840	2.289	3.736	1.933	1.184
PSE: Number valid PSE institutions reported	F3PSENUM	0.975	843	0.032	2.585	1.608	0.020
Current activity-full-time job	F4AACTF	65.658	845	3.099	3.596	1.896	1.635
January 2000 activities same as current	F4AACTD	86.634	824	2.340	3.893	1.973	1.186
Ever served in armed forces	F4AAFTN	97.277	836	0.682	1.467	1.211	0.563
How got current job	F4BHGJ	35.451	818	2.909	3.021	1.738	1.674
Employed for at least 6 months in 1998	F4BJ98A	19.594	797	2.701	3.686	1.920	1.407
Perceived job autonomy	F4BJAUT	40.260	828	3.044	3.186	1.785	1.705
Received license since high school	F4CLICQ	73.414	834	2.729	3.179	1.783	1.531
Training received in last 12 months	F4CTRNQ	44.169	747	2.930	2.598	1.612	1.818
Attended any PSE since high school	F4EANY	44.071	843	2.972	3.017	1.737	1.711
Primarily student or employee	F4ESTEM	81.291	759	2.673	3.561	1.887	1.416
PSE impact-better jobs	F4EJOBS	23.468	764	2.975	3.760	1.939	1.534
Participate in adult education	F4FAEQRY	88.345	832	1.481	1.769	1.330	1.113
Marital status in 2000	F4GMRS	24.564	841	2.618	3.108	1.763	1.485
Total number of dependents	F4GNDP	0.413	839	0.051	3.522	1.877	0.027
Income of respondent in 1999	F4HI99	929.486	763	187.026	6.824	2.612	71.595
Housing status	F4HHOSE	20.377	827	2.494	3.168	1.780	1.402
Registered to vote	F4IRVOTE	65.762	825	2.718	2.704	1.644	1.653
Use of cigarettes	F4ISMOKE	14.065	830	1.724	2.038	1.428	1.207
Hispanic origin	F4JHISP	3.338	829	1.003	2.583	1.607	0.624
Diversity in present neighborhood	F4JRDVB	25.354	799	1.828	2.827	1.681	1.087
Minimum					1.467	1.211	
Maximum					6.824	2.612	
Mean					3.094	1.742	
Standard deviation					0.929	0.247	
Median					3.024	1.739	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-5.—Standard errors and design effects for *Hispanic respondents* to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	39.685	1334	3.203	5.715	2.391	1.340
How many times did R cut/skip classes	F2S9B	6.898	1328	1.297	3.476	1.864	0.696
How far in school mother wants R to go	F2S42B	35.073	1200	2.965	4.630	2.152	1.378
R feels good about him/herself	F2S66A	92.359	1193	1.510	3.854	1.963	0.769
# R's friends to attend 4 yr school	F2S69E	43.244	1165	2.873	3.914	1.978	1.452
Number of jobs during period-#1	NUMJOBS1	1.521	1195	0.043	2.847	1.687	0.026
Time spent working on hobbies	HOBBIES	47.703	1603	2.209	3.132	1.770	1.248
Respondent/family member seriously ill	ILLDISBL	26.093	1600	1.843	2.816	1.678	1.098
Importance of success in work	SUCSLWRK	89.755	1604	1.544	4.158	2.039	0.757
PSE: Number valid PSE institutions reported	F3PSENUM	0.628	1609	0.033	3.855	1.963	0.017
Current activity-full-time job	F4AACTF	74.458	1621	2.279	4.425	2.104	1.083
January 2000 activities same as current	F4AACTD	90.392	1598	1.593	4.668	2.160	0.737
Ever served in armed forces	F4AAFTN	94.099	1589	1.065	3.242	1.801	0.591
How got current job	F4BHGJ	38.937	1563	2.382	3.727	1.931	1.234
Employed for at least 6 months in 1998	F4BJ98A	13.131	1539	1.872	4.727	2.174	0.861
Perceived job autonomy	F4BJAUT	41.403	1589	2.461	3.966	1.991	1.236
Received license since high school	F4CLICQ	77.842	1586	1.974	3.580	1.892	1.043
Training received in last 12 months	F4CTRNQ	49.058	1477	2.745	4.450	2.109	1.301
Attended any PSE since high school	F4EANY	56.391	1618	2.562	4.317	2.078	1.233
Primarily student or employee	F4ESTEM	70.149	1118	2.590	3.578	1.892	1.369
PSE impact-better jobs	F4EJOBS	35.841	1169	3.147	5.031	2.243	1.403
Participate in adult education	F4FAEQRY	86.364	1585	1.611	3.489	1.868	0.862
Marital status in 2000	F4GMRS	40.782	1614	2.527	4.266	2.066	1.224
Total number of dependents	F4GNDP	1.099	1612	0.070	5.301	2.302	0.031
Income of respondent in 1999	F4HI99	833.731	1487	102.483	4.394	2.096	48.889
Housing status	F4HHOSE	26.221	1579	2.061	3.464	1.861	1.107
Registered to vote	F4IRVOTE	71.934	1577	2.594	5.253	2.292	1.132
Use of cigarettes	F4ISMOKE	23.370	1581	2.783	6.833	2.614	1.065
Hispanic origin	F4JHISP	90.021	1579	1.275	2.856	1.690	0.754
Diversity in present neighborhood	F4JRDVB	50.713	1551	2.424	7.265	2.695	0.899
Minimum					2.816	1.678	
Maximum					7.265	2.695	
Mean					4.241	2.045	
Standard deviation					1.067	0.248	
Median					4.062	2.015	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-6.—Standard errors and design effects for *Black, non-Hispanic respondents* to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	18.395	969	3.013	5.856	2.420	1.245
How many times did R cut/skip classes	F2S9B	4.479	974	1.974	8.860	2.977	0.663
How far in school mother wants R to go	F2S42B	32.409	847	4.035	6.289	2.508	1.609
R feels good about him/herself	F2S66A	96.904	876	1.187	4.110	2.027	0.585
# R's friends to attend 4 yr school	F2S69E	48.550	834	5.047	8.498	2.915	1.731
Number of jobs during period-#1	NUMJOBS1	1.425	738	0.056	4.532	2.129	0.026
Time spent working on hobbies	HOBBIES	38.019	1151	2.837	3.929	1.982	1.431
Respondent/family member seriously ill	ILLDISBL	30.137	1149	2.641	3.804	1.950	1.354
Importance of success in work	SUCSLWRK	94.667	1150	1.289	3.783	1.945	0.663
PSE: Number valid PSE institutions reported	F3PSENUM	0.666	1152	0.041	3.915	1.979	0.021
Current activity-full-time job	F4AACTF	73.760	1173	2.674	4.331	2.081	1.285
January 2000 activities same as current	F4AACTD	89.420	1161	1.805	3.996	1.999	0.903
Ever served in armed forces	F4AAFTN	91.061	1151	2.056	5.974	2.444	0.841
How got current job	F4BHGJ	35.224	1133	3.491	6.045	2.459	1.420
Employed for at least 6 months in 1998	F4BJ98A	15.354	1118	2.411	4.998	2.236	1.079
Perceived job autonomy	F4BJAUT	33.073	1156	2.971	4.604	2.146	1.384
Received license since high school	F4CLICQ	73.355	1147	4.450	11.612	3.408	1.306
Training received in last 12 months	F4CTRNQ	50.676	1078	4.149	7.419	2.724	1.523
Attended any PSE since high school	F4EANY	51.968	1170	3.978	7.410	2.722	1.461
Primarily student or employee	F4ESTEM	73.946	821	6.328	17.048	4.129	1.533
PSE impact-better jobs	F4EJOBS	28.966	851	3.755	5.827	2.414	1.556
Participate in adult education	F4FAEQRY	92.370	1140	1.214	2.383	1.544	0.787
Marital status in 2000	F4GMRS	22.130	1164	2.789	5.248	2.291	1.217
Total number of dependents	F4GNDP	1.191	1161	0.099	7.269	2.696	0.037
Income of respondent in 1999	F4HI99	758.134	1023	80.951	2.142	1.464	55.309
Housing status	F4HHOSE	13.232	1135	1.846	3.368	1.835	1.006
Registered to vote	F4IRVOTE	83.568	1139	2.654	5.836	2.416	1.098
Use of cigarettes	F4ISMOKE	19.126	1136	2.644	5.131	2.265	1.167
Hispanic origin	F4JHISP	1.301	1136	0.392	1.355	1.164	0.336
Diversity in present neighborhood	F4JRDVB	55.817	1119	3.497	11.936	3.455	1.012
Minimum					1.355	1.164	
Maximum					17.048	4.129	
Mean					5.917	2.357	
Standard deviation					3.219	0.610	
Median					5.189	2.278	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-7.—Standard errors and design effects for *White, non-Hispanic respondents* to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	11.995	7394	0.620	2.696	1.642	0.378
How many times did R cut/skip classes	F2S9B	4.086	7394	0.500	4.709	2.170	0.230
How far in school mother wants R to go	F2S42B	41.984	7059	0.861	2.149	1.466	0.587
R feels good about him/herself	F2S66A	93.288	6914	0.406	1.816	1.348	0.301
# R's friends to attend 4 yr school	F2S69E	56.841	7016	1.009	2.910	1.706	0.591
Number of jobs during period-#1	NUMJOBS1	1.535	6857	0.015	2.049	1.431	0.011
Time spent working on hobbies	HOBBIES	54.983	8221	0.833	2.305	1.518	0.549
Respondent/family member seriously ill	ILLDISBL	28.004	8213	0.827	2.783	1.668	0.495
Importance of success in work	SUCSLWRK	89.025	8222	0.499	2.099	1.449	0.345
PSE: Number valid PSE institutions reported	F3PSENUM	0.812	8249	0.014	3.526	1.878	0.007
Current activity-full-time job	F4AACTF	78.334	8306	0.702	2.412	1.553	0.452
January 2000 activities same as current	F4AACTD	94.281	8111	0.378	2.149	1.466	0.258
Ever served in armed forces	F4AAFTN	93.520	8194	0.419	2.369	1.539	0.272
How got current job	F4BHGJ	38.857	8123	0.840	2.413	1.553	0.541
Employed for at least 6 months in 1998	F4BJ98A	9.747	8034	0.456	1.901	1.379	0.331
Perceived job autonomy	F4BJAUT	31.061	8239	0.734	2.073	1.440	0.510
Received license since high school	F4CLICQ	70.125	8182	0.826	2.667	1.633	0.506
Training received in last 12 months	F4CTRNQ	40.898	7732	0.880	2.475	1.573	0.559
Attended any PSE since high school	F4EANY	55.343	8293	0.905	2.750	1.658	0.546
Primarily student or employee	F4ESTEM	78.771	6319	0.833	2.622	1.619	0.514
PSE impact-better jobs	F4EJOBS	23.082	6496	0.870	2.772	1.665	0.523
Participate in adult education	F4FAEQRY	88.192	8170	0.575	2.591	1.610	0.357
Marital status in 2000	F4GMRS	43.474	8286	0.955	3.076	1.754	0.545
Total number of dependents	F4GNDP	0.632	8280	0.022	4.334	2.082	0.011
Income of respondent in 1999	F4HI99	715.642	7698	22.828	1.556	1.248	18.298
Housing status	F4HHOSE	33.969	8156	0.896	2.918	1.708	0.524
Registered to vote	F4IRVOTE	78.667	8126	0.778	2.934	1.713	0.454
Use of cigarettes	F4ISMOKE	29.666	8151	0.860	2.892	1.700	0.506
Hispanic origin	F4JHISP	2.336	8153	0.354	4.490	2.119	0.167
Diversity in present neighborhood	F4JRDVB	73.115	8022	0.494	2.942	1.715	0.288
Minimum					1.556	1.248	
Maximum					4.709	2.170	
Mean					2.713	1.633	
Standard deviation					0.743	0.214	
Median					2.645	1.626	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-8.—Standard errors and design effects for *American Indian and Alaska Native* respondents to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[1]	DEFF	DEFT	SRS SE[2]
There are many gangs in school	F2S7H	28.188	107	5.862	1.804	1.343	4.365
How many times did R cut/skip classes	F2S9B	0.738	107	0.696	0.703	0.839	0.830
How far in school mother wants R to go	F2S42B	17.365	97	3.878	1.009	1.005	3.860
R feels good about him/herself	F2S66A	90.999	100	4.159	2.098	1.448	2.872
# R's friends to attend 4 yr school	F2S69E	36.002	100	5.816	1.458	1.208	4.817
Number of jobs during period-#1	NUMJOBS1	1.430	97	0.103	1.905	1.380	0.075
Time spent working on hobbies	HOBBIES	55.275	144	8.283	3.969	1.992	4.158
Respondent/family member seriously ill	ILLDISBL	41.057	143	9.542	5.343	2.311	4.128
Importance of success in work	SUCSLWRK	80.847	144	9.445	8.239	2.870	3.291
PSE: Number valid PSE institutions reported	F3PSENUM	0.369	145	0.073	2.008	1.417	0.052
Current activity-full-time job	F4AACTF	61.559	145	13.035	10.339	3.215	4.054
January 2000 activities same as current	F4AACTD	91.919	143	2.752	1.448	1.203	2.287
Ever served in armed forces	F4AAFTN	94.699	141	2.419	1.632	1.278	1.893
How got current job	F4BHGJ	36.858	134	7.004	2.805	1.675	4.182
Employed for at least 6 months in 1998	F4BJ98A	11.429	132	3.903	1.972	1.404	2.779
Perceived job autonomy	F4BJAUT	41.859	139	7.740	3.398	1.843	4.199
Received license since high school	F4CLICQ	75.959	140	6.112	2.844	1.686	3.624
Training received in last 12 months	F4CTRNQ	59.465	128	8.890	4.168	2.042	4.355
Attended any PSE since high school	F4EANY	72.139	144	7.385	3.880	1.970	3.749
Primarily student or employee	F4ESTEM	77.399	79	6.182	1.714	1.309	4.722
PSE impact-better jobs	F4EJOBS	28.983	81	7.025	1.929	1.389	5.058
Participate in adult education	F4FAEQRY	94.512	140	2.614	1.831	1.353	1.932
Marital status in 2000	F4GMRS	39.019	144	8.875	4.734	2.176	4.079
Total number of dependents	F4GNDP	1.457	144	0.153	2.262	1.504	0.102
Income of respondent in 1999	F4HI99	1782.361	128	588.405	5.229	2.287	257.306
Housing status	F4HHOSE	41.261	139	7.094	2.867	1.693	4.190
Registered to vote	F4IRVOTE	76.644	139	5.930	2.711	1.647	3.601
Use of cigarettes	F4ISMOKE	24.671	139	6.679	3.313	1.820	3.669
Hispanic origin	F4JHISP	18.682	140	7.833	5.616	2.370	3.306
Diversity in present neighborhood	F4JRDVB	66.031	138	7.648	5.997	2.449	3.123
Minimum					0.703	0.839	
Maximum					10.339	3.215	
Mean					3.308	1.738	
Standard deviation					2.176	0.546	
Median					2.758	1.661	

[1] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[2] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-9.—Standard errors and design effects for *dropout respondents*[1] to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[2]	DEFF	DEFT	SRS SE[3]
There are many gangs in school	F2S7H	27.330	276	3.948	2.165	1.471	2.683
How many times did R cut/skip classes	F2S9B	18.377	274	3.839	2.691	1.640	2.341
How far in school mother wants R to go	F2S42B	41.399	255	5.128	2.763	1.662	3.085
R feels good about him/herself	F2S66A	96.800	229	1.087	0.872	0.934	1.163
# R's friends to attend 4 yr school	F2S69E	27.369	263	3.775	1.884	1.373	2.750
Number of jobs during period-#1	NUMJOBS1	1.491	1049	0.048	3.138	1.771	0.027
Time spent working on hobbies	HOBBIES	44.617	1519	2.239	3.079	1.755	1.276
Respondent/family member seriously ill	ILLDISBL	34.238	1519	2.294	3.549	1.884	1.218
Importance of success in work	SUCSLWRK	87.475	1519	1.596	3.531	1.879	0.850
PSE: Number valid PSE institutions reported	F3PSENUM	0.260	1529	0.023	3.094	1.759	0.013
Current activity-full-time job	F4AACTF	68.180	1550	2.310	3.810	1.952	1.183
January 2000 activities same as current	F4AACTD	90.477	1527	1.136	2.287	1.512	0.751
Ever served in armed forces	F4AAFTN	97.392	1510	0.560	1.865	1.366	0.410
How got current job	F4BHGJ	47.226	1458	2.450	3.510	1.873	1.308
Employed for at least 6 months in 1998	F4BJ98A	18.059	1420	1.684	2.719	1.649	1.021
Perceived job autonomy	F4BJAUT	32.648	1493	2.098	2.988	1.729	1.214
Received license since high school	F4CLICQ	77.798	1509	1.727	2.604	1.614	1.070
Training received in last 12 months	F4CTRNQ	60.742	1334	2.457	3.374	1.837	1.337
Attended any PSE since high school	F4EANY	68.465	1543	2.295	3.762	1.940	1.183
Primarily student or employee	F4ESTEM	69.187	578	3.379	3.094	1.759	1.921
PSE impact-better jobs	F4EJOBS	38.074	605	3.545	3.222	1.795	1.975
Participate in adult education	F4FAEQRY	89.909	1504	1.446	3.466	1.862	0.777
Marital status in 2000	F4GMRS	40.825	1542	2.289	3.342	1.828	1.252
Total number of dependents	F4GNDP	1.563	1543	0.067	3.963	1.991	0.034
Income of respondent in 1999	F4HI99	975.965	1359	95.515	2.794	1.672	57.140
Housing status	F4HHOSE	32.573	1498	2.059	2.890	1.700	1.211
Registered to vote	F4IRVOTE	70.976	1487	1.941	2.719	1.649	1.177
Use of cigarettes	F4ISMOKE	48.519	1500	2.251	3.040	1.744	1.291
Hispanic origin	F4JHISP	19.161	1498	2.238	4.840	2.200	1.017
Diversity in present neighborhood	F4JRDVB	65.124	1466	1.588	3.637	1.907	0.833
Minimum					0.872	0.934	
Maximum					4.840	2.200	
Mean					3.023	1.724	
Standard deviation					0.746	0.233	
Median					3.086	1.757	

[1] This table is based on NELS:88/2000 respondents who had "ever dropped out" of school during the first and second follow-up studies.

[2] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[3] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Table D-10.—Standard errors and design effects for *non-dropout respondents*[1] to the NELS:88 fourth follow-up study

Label	Variable	Estimate	N	Design SE[2]	DEFF	DEFT	SRS SE[3]
There are many gangs in school	F2S7H	15.976	10360	0.801	4.949	2.225	0.360
How many times did R cut/skip classes	F2S9B	4.044	10369	0.472	5.947	2.439	0.193
How far in school mother wants R to go	F2S42B	39.713	9735	0.890	3.217	1.794	0.496
R feels good about him/herself	F2S66A	93.422	9636	0.380	2.264	1.505	0.253
# R's friends to attend 4 yr school	F2S69E	55.445	9630	1.050	4.294	2.072	0.507
Number of jobs during period-#1	NUMJOBS1	1.520	8437	0.014	2.089	1.445	0.009
Time spent working on hobbies	HOBBIES	53.574	10494	0.775	2.537	1.593	0.487
Respondent/family member seriously ill	ILLDISBL	26.630	10477	0.713	2.726	1.651	0.432
Importance of success in work	SUCSLWRK	90.153	10494	0.453	2.429	1.558	0.291
PSE: Number valid PSE institutions reported	F3PSENUM	0.881	10522	0.012	3.462	1.861	0.006
Current activity-full-time job	F4AACTF	78.468	10593	0.666	2.782	1.668	0.399
January 2000 activities same as current	F4AACTD	93.273	10362	0.450	3.346	1.829	0.246
Ever served in armed forces	F4AAFTN	92.645	10454	0.489	3.667	1.915	0.255
How got current job	F4BHGJ	36.520	10366	0.863	3.331	1.825	0.473
Employed for at least 6 months in 1998	F4BJ98A	9.870	10250	0.489	2.751	1.659	0.295
Perceived job autonomy	F4BJAUT	33.229	10511	0.768	2.794	1.671	0.459
Received license since high school	F4CLICQ	70.279	10433	0.924	4.264	2.065	0.447
Training received in last 12 months	F4CTRNQ	39.965	9876	0.864	3.074	1.753	0.493
Attended any PSE since high school	F4EANY	52.031	10578	0.935	3.703	1.924	0.486
Primarily student or employee	F4ESTEM	78.374	8548	1.083	5.920	2.433	0.445
PSE impact-better jobs	F4EJOBS	24.028	8788	0.892	3.832	1.958	0.456
Participate in adult education	F4FAEQRY	88.187	10416	0.506	2.559	1.600	0.316
Marital status in 2000	F4GMRS	39.427	10559	0.936	3.870	1.967	0.476
Total number of dependents	F4GNDP	0.591	10546	0.020	4.893	2.212	0.009
Income of respondent in 1999	F4HI99	710.324	9787	23.609	2.122	1.457	16.208
Housing status	F4HHOSE	29.712	10390	0.854	3.630	1.905	0.448
Registered to vote	F4IRVOTE	79.462	10372	0.771	3.777	1.944	0.397
Use of cigarettes	F4ISMOKE	22.203	10390	0.750	3.379	1.838	0.408
Hispanic origin	F4JHISP	10.934	10392	0.860	7.883	2.808	0.306
Diversity in present neighborhood	F4JRDVB	66.642	10214	0.810	7.136	2.671	0.303
Minimum					2.089	1.445	
Maximum					7.883	2.808	
Mean					3.754	1.908	
Standard deviation					1.425	0.342	
Median					3.421	1.849	

[1] This table is based on the responses for NELS:88/2000 respondents who never reported dropping out of school.

[2] Standard errors calculated under with-replacement clustered sampling design using SUDAAN software.

[3] Standard errors calculated under simple random sampling design.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988 (NELS:88), 2000.

Appendix E: Glossary

Appendix E
Glossary

Adaptive (Two-Stage Multilevel) Testing: After the NELS:88 base-year, multiple test forms of varying levels of difficulty were assigned based on the prior round ability estimate (theta) in the subject. Thus the specific sequence of questions each student answered was tailored to that student's ability level. This procedure helps to minimize floor and ceiling effects (see entries for "Ceiling effect" and "Floor effect").

Alternative completer: The NELS:88 second follow-up distinguished three levels of enrollment status: students enrolled in a regular high school program, dropouts who had enrolled in (or had completed) some alternative (non-diploma) high school equivalency accrediting program (for example, preparation classes for the GED test), and dropouts receiving no alternative instruction. The term "alternative completer" was used for dropouts receiving any sort of instruction to prepare them for equivalency certification, and for dropouts who had already received the GED or other equivalency certification. In terms of questionnaire completion, alternative completers were treated in two ways. Dropouts receiving alternative instruction in preparation for possible equivalency certification were administered the dropout questionnaire. Those dropouts who had received the GED or other high school equivalency certification were treated as school completers, and were administered the student questionnaire.

ASCII: American Standard Code for Information Interchange. A standard method for encoding characters; includes codes representing upper and lower case letters, numerals, and punctuation.

Base Year Ineligibles (BYI) Study: A NELS:88 first and second follow-up study which sought to locate and survey eligible respondents who were part of the Base-year sample, yet were ineligible to participate in the Base Year due to mental or physical disability or to a language barrier such as limited or no proficiency in English.

Bayesian statistics: Bayesian methods incorporate the prior probability distribution with the new evidence collected, as was done in rescaling NELS:88 1988 to 1992 test results when the 1992 test data became available.

Bias: Bias is the difference between the reported value and the true value. Thus the bias of an estimate is the difference between the expected value of a sample estimate and the corresponding true value for the population. Response bias is the difference between respondent reports and their behavior or characteristics. Nonresponse bias is the difference that occurs when respondents differ as a group from nonrespondents on a characteristic being studied. Sample bias is the unequal selection or the omission of members of the population, without appropriate weighting. Relatedly, undercoverage bias arises because some portion of the potential sampling frame is missed or excluded, or there are duplicate units. For example, if the school list from which a school sample is drawn is incomplete or inaccurate (owing, for example, to the birth of new schools subsequent to the time the list was drawn up), school undercoverage may occur.

Burden: Formally, this is the aggregate hours realistically required for data providers to participate in a data collection. Burden also has a subjective or psychological dimension: the degree to which providing information is regarded as onerous may depend on the salience to the respondent of the questions that are being posed and on other factors such as competing time demands.

BY: NELS:88 Base Year Study of 8th-graders conducted in 1988.

BYI: See "Base Year Ineligibles Study."

Carnegie units: A standard of measurement used for secondary education that represents the completion of a course that meets one period per day for one year.

CAPI (Computer-assisted personal interviewing): CAPI in which the questionnaire is loaded into a field interviewer's laptop computer. CAPI was used to complete some cases in the NELS:88 fourth follow-up (2000) and incorporated the same instrument that was used with CATI.

CATI (Computer-assisted telephone interviewing): CATI was the dominant mode of data collection in the NELS:88 third (1994) and fourth (2000) follow-ups.

CCD (Common Core of Data): Data annually collected from all public schools in the United States by the National Center for Education Statistics.

CD-ROM: NELS:88 data are distributed primarily in an optical laser disc medium, specifically, CD-ROM (Compact Disc Read-Only Memory). A CD-ROM is a computer storage disc in the same physical form as an audio CD and can store approximately 650 megabytes of digital data.

Ceiling effect: The result of a cognitive test having insufficient numbers of the more difficult items. In a longitudinal study, ceiling effects in the follow-up testings can cause change scores to be artificially constrained for high ability examinees. More information (that is, smaller error of measurement) is obtained with respect to ability level if high ability individuals receive relatively harder items (and if low ability individuals receive proportionately easier items). The matching of item difficulty to a person's ability level yields increased reliability at the extremes of the score distribution where it is most needed for studies of longitudinal change. That is, the measurement problems related to floor and ceiling effects in combination with regression effects found at the extreme score ranges seriously hamper the accuracy of change measures in longitudinal studies. Hence one strategy employed in NELS:88 to minimize ceiling effects was to develop test forms that are "adaptive" to the ability level of the examinee. The multilevel tests used in the first and second follow-ups of NELS:88--with test assignment based on prior test performance--work to minimize the possibility of ceiling effects biasing the estimates of the score gains. (See entry for "Floor effect.")

Certainty school: A first or second follow-up school attended by four or more NELS:88 sample members, as determined by tracing and data collection efforts. These schools are included in the sample with certainty (probability = 1). All NELS:88 first follow-up sample members in the school at the time of data collection were included in the second follow-up.

Closed-ended: A type of question in which the data provider's responses are limited to given alternatives (as opposed to an open-ended question. See entry for "Open-ended.")

Cluster size: The number of NELS:88 sample members attending a particular high school.

Clustering: A sample selection method in which small geographical areas such as schools (e.g. in NELS:88), school districts, counties, or blocks are selected as an initial stage, with individuals selected in a subsequent step. (See entry for "Primary Sampling Unit.")

Codebook: A record of each variable being measured, including variable name, columns occupied by each variable in the data matrix, values used to define each variable, unweighted frequencies, unweighted percents, and weighted valid percents. (See entry for "Electronic codebook.")

Cognitive test battery: One of the two parts of the Student Survey (the second part being the student questionnaire). Four achievement areas (mathematics, reading, science, and social studies [history/ citizenship/geography]) were measured.

Cohort: A group of individuals who have a statistical factor in common, for example, year of birth or grade in school or year of high school graduation. NELS:88 embraces three overlapping but distinct nationally representative grade cohorts: spring-term 1988 8th-graders, 1990 high school sophomores, and 1992 high school seniors.

Composite variables: A composite variable is one that is constructed through either the combination of two or more variables (socioeconomic status, for example) or calculated through the application of a mathematical function to a variable. Also called a "derived variable" or "constructed variable."

Confidence interval: A sample-based estimate expressed as an interval or range of values within which the true population value is expected to be located (with a specified degree of confidence).

Contextual data: In NELS:88, the primary unit of analysis is the student (or dropout), and information from the other study components, referred to as the contextual data, should be viewed as extensions of the student data--for example, as school administrator, teacher, and parent reports on the student's school learning environment or home situation.

Core school: School that was selected between phases 1 and 2 of the Second Follow-up to receive the full complement (School Administrator, Teacher, Transcript) of study components, and for in-school data collection sessions.

Core student: Students who are part of the primary cohort of NELS:88, in contrast to state augmentation or School Effectiveness Study students. The core students include those chosen as 8th-graders in the 1988 Base Year Study and those added to the sample through freshening procedures during the first or second follow-up.

Core study: The original NELS:88 study, in contrast to the study with additions and follow-up additions like the state augmentation studies and the School Effectiveness Study.

Course offerings: School-level summaries of courses offered and of course enrollment levels; while in HS&B course offerings data were collected for all schools, in NELS:88 such data have been collected only for schools in the High School Effectiveness Study.

Cross-sectional survey: A cross-sectional design represents events and statuses at a single point in time. For example, a cross-sectional survey may measure the cumulative educational attainment (achievements, attitudes, statuses) of students at a particular stage of schooling (for example, 8th grade, 10th grade, or 12th grade). In contrast, a longitudinal (or repeated measurement of the same sample units) survey measures the change or growth in educational attainments that occurs over a particular period of schooling. The longitudinal design of NELS:88 generates--by means of sample "freshening"--three representative cross-sections (8th-graders in 1988, high school sophomores in 1990, seniors in 1992) and permits analysis of individual level change over time through longitudinal analysis and of group level and intercohort change through the cross-sectional comparisons. NELS:88 provides analysts with both cross-sectional weights, and panel (longitudinal) weights, so that both status at one point in time, and growth or change over time, can be investigated. (See entry for "Longitudinal or panel survey.")

Data element: The most basic unit of information. In data processing it is the fundamental data structure. It is defined by its size (in characters) and data type (e.g., alphanumeric, numeric only, true/false, date) and may include a specific set of values or range of values.

Design effect: A measure of sample efficiency. The design effect (DEFF) is the variance of an estimate divided by the variance of the estimate that would have occurred if a sample of the same size had been selected using simple random sampling. Sometimes it is more useful to work with standard errors than with variances. The root design effect (DEFT) expresses the relation between the actual standard error of an estimate and the standard error of the corresponding estimates from a simple random sample.

Design weights: Design weights compensate for unequal probabilities of selection. They are also called raw weights, unadjusted weights, or sampling weights. Design weights may be contrasted to adjusted weights (adjusted to compensate for nonresponse, and also called final weights or analysis weights). Roughly, the design weight is calculated as the inverse of the probability of selection, taking into account all stages of the sample selection process. More precisely, design weights are the inverses of the expected frequencies with which population units appear in conceptually repeated samples selected using the sampling design developed for the study. Design weights do not appear on the NELS:88 data files, since they are typically not used in analysis. (See also the entry for Nonresponse-adjusted weights.)

Dropout: The term is used both to describe an event (leaving school before graduating) and a status (an individual who is not in school and is not a graduate at a defined point in time). The "cohort dropout rate" in NELS:88 is based on measurement of enrollment status of 1988 8th-graders two and four years later (that is, in the spring term of 1990 and the spring term of 1992) and of 1990 sophomores two years later. A respondent who has not graduated from high school or attained an equivalency certificate and who has not attended high school for 20 consecutive days (not counting any excused absences) is considered to be a dropout. In contrast, transferring schools--for example, from a public to a private school--is not regarded as a dropout event, nor is delayed graduation (as when a student is continuously enrolled but takes an additional year to complete school). A person who drops out of school may later return and graduate: at the time the person left school initially, he or she is called a "dropout," and at the time the person returns to school, he or she is called a "stopout."

Early graduate: A student who graduated from high school in less than the typical amount of time. For example, if a student graduated in December of his/her senior year (when the majority of his/her classmates graduate the following May or June), the student is categorized as an early graduate. Some dropouts became exam-certified for high school equivalency (e.g., by taking the GED) ahead of the scheduled graduation date for their original grade cohort—these individuals, too, are classified as early graduates. In the main study data collection, early graduates were administered a special supplement in the student questionnaire along with the cognitive test battery.

Electronic codebook (ECB): While hardcopy codebooks with item stems, response categories, associated response frequency distributions, unweighted percents, and weighted valid percents are contained within the NELS:88 base year through second follow-up user's manuals, NELS:88 data are also available on CD-ROM in an electronic codebook (ECB) format. Electronic codebooks are menu-driven systems that allows users to perform functions such as the following: (a) search a list of database variables based upon key words or variable names/labels; (b) display weighted and unweighted percentages for each variable in the database; (c) display question text for each variable in the database; (d) select or tag variables for subsequent analysis; (e) generate SAS-PC or SPSS-PC+ program code/command statements for subsequently constructing a system file of the selected variables; and (f) generate a codebook of the selected variables. Three electronic codebooks have been prepared for public use, and three parallel ECBs for restricted use: base year through second follow-up (1988-92)

containing all sample cases who participated in any of the in-school rounds; base year through third follow-up, based on the 1994 sample; and base year through fourth follow-up, based on the 2000 sample.

Equated test score: Test equating takes place in two distinct contexts in NELS:88. One context is *vertical equating* of forms for use in successive grades such that the achievement growth of individual NELS:88 sample members over time can be accurately measured. Another context is *cross-sectional equating.* There are two examples of such cross-sectional linkage of tests: first, the NELS:88 mathematics test was put on the HS&B scale to facilitate cross-cohort or trend comparisons; second, the NELS:88 1992 mathematics test was put on the NAEP scale.

ETS: Educational Testing Service. The National Opinion Research Center's (NORC's) subcontractor for NELS:88 cognitive test development and evaluation. (See entry for "NORC.")

Expanded sample: the combined sample of eligible and ineligible NELS:88 sample members, including 8th-graders who were excluded from the survey. This sample can be used to make unbiased estimates of national dropout rates.

F1: The NELS:88 first follow-up, conducted in 1990; sometimes used as a variable prefix.

F2: The NELS:88 second follow-up, conducted in 1992; sometimes used as a variable prefix.

F3: The NELS:88 third follow-up (NELS:88/94) conducted in 1994; sometimes used as a variable prefix.

F4: The NELS:88 fourth follow-up (NELS:88/2000) conducted in 2000; sometimes used as a variable prefix.

File: Refers to a data file containing a set of related computerized records.

Floor effect: The result of a cognitive test being too difficult for a large number of the examinees, causing the low ability examinees to receive chance scores on the first testing, and on subsequent testings if the test remains too difficult. Floor effects result in an inability to discriminate among low ability individuals at time one or time two, and there will be no reliable discrimination among examinees with respect to amounts of change. A possible solution, utilized in NELS:88, is to develop test forms that are "adaptive" to the ability level of the examinee, which tends to minimize the possibility of floor effects biasing the estimates of the score gains.

Freshening: A NELS:88 sampling procedure by which high school sophomores were added in the first follow-up who were not in the 8th grade in the United States 2 years before. This process was repeated in the second follow-up, adding high school seniors who were not in the 8th grade in the U.S. 4 years before, and not in the 10th grade in the United States 2 years before. This process ensured that the sample would be representative of the 1992 senior class by allowing 1992 seniors who did not have a chance for selection into the base-year (or the first follow-up) sample to have some probability of 1992 selection.

GED recipient: A person who has obtained certification of high school equivalency by meeting state requirements and passing an approved exam, which is intended to provide an appraisal of the person's achievement or performance in the broad subject matter areas usually required for high school graduation. (See entries for "GED test" and "Alternative completer.")

GED test: General Educational Development test. A test administered by the American Council on Education as the basis for awarding a high school equivalent certification.

HS&B (High School and Beyond): The second in the series of longitudinal education studies sponsored by NCES. The HS&B Base Year study surveyed sophomore and senior students in 1980.

High School Effectiveness Study (HSES): The NELS:88 High School Effectiveness Study (HSES) is a special component of NELS:88 that was designed to estimate school-level characteristics. HSES consists of a sample of 247 urban and suburban 10th grade schools in the 30 largest metropolitan statistical areas (MSAs). For comparison purposes, HSES used eight basic strata defined on the basis of four types of schools (Public, Catholic, NAIS, and Other Private) at two levels of urbanicity (Urban, Suburban). HSES substantially increased cluster sizes and provided in-school representative student samples; selection probabilities were simulated for the schools so that school weights could be generated. This component was continued in the second follow-up, and included student, school administrator, teacher, and parent questionnaires, transcript and course offerings surveys.

IEP: Individualized Education Program in special education for students with a mental or physical disability.

IPEDS (Integrated Postsecondary Education Data System): IPEDS is an NCES postsecondary education data collection program designed to encompass all institutions whose primary purpose is to provide postsecondary education. IPEDS institution codes appear in the NELS:88 restricted-use files (but never on the public files), thus providing further linkages of information about the postsecondary institutions attended by NELS:88 sample members.

IRT: Item Response Theory. A method of estimating achievement level by considering the pattern of right, wrong, and omitted responses on all items administered to an individual student. Rather than merely counting right and wrong responses, the IRT procedure also considers characteristics of each of the test items, such as their difficulty, and the likelihood that they could be guessed correctly by low-ability individuals. IRT scores are less likely than simple number-right or formula scores to be distorted by correct guesses on difficult items if a student's response vector also contains incorrect answers to easier questions. Another attribute of IRT that makes it useful for NELS:88 is the calibration of item parameters for all items administered to all students. This makes it possible to obtain scores on the same scale for students who took harder or easier forms of the test. IRT also permits vertical scaling of the three grade levels (grade 8 in 1988, grade 10 in 1990, grade 12 in 1992).

Item nonresponse: The amount of missing information when a valid response to an item or variable was expected. (See entry for "Unit nonresponse.")

LEP: Limited English Proficient. A concept developed to assist in identifying those language-minority students (individuals from non-English language backgrounds) who need language assistance services, in their own language or in English, in the schools. (See entries for "NEP" and "LM.") The Bilingual Education Act, reauthorized in 1988 (P.L. 100-297), describes a limited-English-proficient student as one who either meets one or more of the following conditions:

a) the student was born outside of the United States or the student's native language is not English;

b) the student comes from an environment where a language other than English is dominant; or

c) the student is American Indian or Alaskan Native and comes from an environment where a language other than English has had a significant impact on his/her level of English language proficiency; or who has sufficient difficulty speaking, reading, writing, or understanding the

English language to deny him or her the opportunity to learn successfully in English-only classrooms.

LM: Language Minority. A non, limited- or fully English proficient student in whose home a non-English language is typically spoken.

Longitudinal or panel survey: In a longitudinal design, similar measurements--of the same sample of individuals, institutions, households or of some other defined unit--are taken at multiple time points. NELS:88 employs a longitudinal design that follows the same individuals over time, and permits the analysis of individual-level change. (See entry for "Cross-sectional survey.")

Machine editing: Also called forced data cleaning or logical editing. Uses computerized instructions in the data cleaning program that ensure common sense consistency within and across the responses from a data provider.

Microdata (microrecords): Observations of individual sample members, such as those contained on the NELS:88 data files.

MSA: Metropolitan statistical area. A large population nucleus and the nearby communities which have a high degree of economic and social integration with that nucleus. Each MSA consists of one or more entire counties (or county equivalents) that meet specified standards pertaining to population, commuting ties, and metropolitan character. (However, in New England, towns and cities, rather than counties, are the basic units.) MSAs are designated by the Office of Management and Budget (OMB). An MSA includes a city and, generally, its entire urban area and the remainder of the county or counties in which the urban area is located. A MSA also includes such additional outlying counties which meet specified criteria relating to metropolitan character and level of community of workers into the central city or counties.

Multidimensional raking: An adjustment procedure in weighting whereby the sum of the weights for each marginal category of respondents in the follow-up rounds of NELS:88 was made equal to the corresponding sum of the final prior round weights for that group. (See also entry for "Raking.")

NAEP: The National Assessment of Educational Progress. NAEP is a cross-sectional assessment program that measures achievement at the group level for students in fourth, 8^{th} and 12^{th} grades and provides a time series for measuring trends in academic progress of 9-, 13- and 17-year olds. NELS:88 tests differ from but complement those of NAEP by providing a basis for measuring individual-level achievement growth between 8^{th}, 10^{th} and 12^{th} grades in mathematics, science, reading and social studies, and relating cognitive gains in these subjects to the individual, school and family factors that are measured in the various NELS:88 questionnaires and school records (transcript) studies.

NAIS: The National Association of Independent Schools. This organization endorsed NELS:88. NAIS schools form a base-year school sampling stratum in NELS:88, and NAIS constitutes a category within the restricted-use file school control type variable.

NCEA: The National Catholic Educational Association. This organization endorsed NELS:88; Catholic schools constituted a separate sampling stratum and are available as a distinct analytic category in NELS:88.

NCES: The National Center for Education Statistics, Office of Educational Research and Improvement, of the U.S. Department of Education. This governmental agency is the primary sponsor of NELS:88, and is also the sponsoring agency for (among other studies) NAEP, ELS:2002, HS&B, and NLS-72.

NELS:88: The National Education Longitudinal Study of 1988. Third in the series of longitudinal education studies sponsored by NCES. The study began in 1988 with the 8th-grade class of that year. The study collected data in 1988, 1990, and 1992 on student's school experiences, as well as background information from school administrators, teachers and parents (in the base-year and second follow-up only). Data on out-of-school experiences were collected in 1994 and 2000.

NEP: No English Proficiency. A student who does not speak English. (See entry for "LEP.")

New Basics: In its report *A Nation At Risk: The Imperative for Educational Reform* (1983), the National Commission on Excellence in Education recommended that all high school students "be required to lay the foundations in the Five New Basics by taking the following curriculum during their four years of high school: (i) 4 years of English; (ii) 3 years of mathematics; (iii) 3 years of science; (iv) 3 years of social studies; and (v) one-half year of computer science." A more stringent version of the New Basics was offered by Secretary of Education William Bennett in 1988 (*American Education, Making It Work: A Report to the President and the American People)*, comprising the scheme above, plus a minimum of two years of foreign language. Summary composite variables, reflecting various interpretations of the New Basics, were created for the HS&B and NAEP high school transcript studies; the NELS:88 transcript study provides both HS&B and NAEP equivalent New Basics variables.

NLS-72: The National Longitudinal Study of the High School Class of 1972. This project was the first in the series of longitudinal education studies sponsored by NCES.

Noncertainty schools: Schools in which fewer than four (three, two or one) NELS:88 students attended. These schools were not subsampled for participation in the School Administrator, Teacher, and Transcript components. Additionally, the survey instruments were not administered in group sessions in the schools, as was done in the certainty schools.

Nonresponse: (See entries for "Item nonresponse" and "Unit nonresponse.")

Nonresponse-adjusted weights: Nonresponse-adjusted weights are sometimes called adjusted weights, final weights, or analysis weights. Building on the design weight, they compensate for nonresponse. In the first four rounds of NELS:88, a weighting cell approach was used for nonresponse adjustment. In the fifth (2000) round, a generalized exponential model was used to unify the approach to nonresponse and post-stratification adjustment. (See entry for "Design Weight.")

Nonsampling error: An error in sample estimates that cannot be attributed to sampling fluctuations. Such errors may arise from many sources including imperfect implementation of sampling procedures, differential unit or item nonresponse across subgroups, bias in estimation, or errors in observation and recording.

NORC: The National Opinion Research Center at The University of Chicago. NORC conducted the first four rounds of NELS:88 for the National Center for Education Statistics.

NSF: The National Science Foundation, one of the sponsors of the in-school rounds of NELS:88. NSF sponsored several components of NELS:88: 1) additions to the student questionnaire to learn about students' experiences and their exposure to mathematics and science curricula; 2) a survey of mathematics and science teachers to obtain evaluations of their NELS:88 student(s) and to learn about their classroom practices and background preparation for teaching; (3) a base-year study of the postsecondary education transcripts of NELS:88 math and science teachers; (4) use of experimental constructed response format math and science achievement test items in the 1992 High School Effectiveness Study schools; and (5) a

validity study in a small subset of NELS:88 second follow-up high schools centering on teacher reports of instructional content, strategies and goals.

OBEMLA: The Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education. OBEMLA funded a NELS:88 supplement in the base year through second follow-up that inquired into the education experiences of students whose native language was not English. The name of this office has now changed to the Office of English Language Acquisition.

OMB: The Office of Management and Budget. OMB is a federal agency with the responsibility for reviewing all studies funded by executive branch agencies. OMB reviewed, commented on, and approved the NELS:88 questionnaires, as indicated by their approval number and its expiration date in the top right corner of the questionnaire covers.

Open-ended: A type of question in which the data provider's responses are not limited to given alternatives.

Optical disc: A disc that is read optically (e.g., by laser technology), rather than magnetically. (See entry for "CD-ROM.")

Optical scanning: A system of recording responses that transfers responses into machine-readable data through optical mark reading. This method of data capture was used for the NELS:88 student questionnaires and cognitive tests, as well as for the parent and teacher questionnaires. (In contrast, responses to certain other questionnaires, such as the school administrator questionnaire, were keyed by using conventional data entry methods.)

Out-of-sequence: Students in a NELS:88 follow-up who are not in the modal grade sequence of their defining cohort. (In other words, the student is not in the grade that he/she would be in if progressing with the majority of the cohort through school.) For example, most NELS:88 spring term 1988 8^{th}-graders were in the 10^{th} grade in the 1989-90 school year; one would be described as out-of-sequence if that student had skipped a grade and was in the eleventh grade in the 1989-90 school year, or had been held back and was in the ninth grade, or had dropped out and was in not in school at all.

Parent, NELS-targeted parent/guardian: The NELS:88 Parent Component sought to collect information from parents of eligible student/dropout respondents. It was asked that the parent or guardian who knew most about his or her child's educational experience complete the questionnaire.

PETS: The Postsecondary Education Transcript Study conducted by RTI for NCES as part of the fourth follow-up (NELS:88/2000). Postsecondary transcripts were also collected in NLS-72 and HS&B. (See also entry for "RTI.")

PIN: Personal Identification Number. In NELS:88, a unique number assigned to each district and school.

Population: All individuals in the group to which conclusions from a data collection activity are to be applied. Weighted results of NELS:88 data provide estimates for populations and subgroups.

Population variance: A measure of dispersion defined as the average of the squared deviations between the observed values of the elements of a population or sample and the population mean of those values.

Postsecondary education: The provision of formal instructional programs with a curriculum designed primarily for students who have completed the requirements for a high school diploma or equivalent.

This includes programs of an academic, vocational, and continuing professional education purpose, and excludes avocational and adult basic education programs.

Poststratification adjustment: A weight adjustment that forces survey estimates to match independent population totals within selected poststrata (adjustment cells).

Precision: The difference between a sample-based estimate and its expected value. Precision is measured by the sampling error (or standard error) of an estimate.

Primary Sampling Unit (PSU): Unit chosen at the first stage of a cluster sample. In NELS:88, the PSU is the school; in other studies, geographical units such as a county or MSA may serve as the PSU.

Probability sample: A sample selected by a method such that each unit has a fixed and determined probability of selection -- i.e., each population unit has a known, nonzero chance of being included.

QED: Quality Education Data. QED is a commercial firm that publishes national directories of all public and private schools and districts. Its list of schools in the U.S. constituted the sampling frame for the base year, and provided important information on school location, principal's name, minority enrollment, and other characteristics.

Raking: Raking is a general term for statistical methods used to adjust sampling weights in survey data. In NELS:88, for example, the approach was used for nonresponse weight adjustment and weight calibration. The general purpose of raking in NELS:88 was to ensure that when analysis was performed on successive waves, there would not be changes in demographics or dropout rates that were an artifact of lack of control in weights. In the fourth follow-up study, the GEM model, a generalization of the raking technique, minimally adjusted sampling weights for survey nonresponse while maintaining pre-specified group totals for the study population. (See also entry for "Multidimensional Raking.")

Range check: A determination of whether responses fall within a predetermined set of acceptable values.

Record format: The layout of the information contained in a data record (includes the name, type, and size of each field in the record).

Records: A logical grouping of data elements within a file upon which a computer program acts.

Reliability: The consistency in results of a test or measurement including the tendency of the test or measurement to produce the same results when applied twice to some entity or attribute believed not to have changed in the interval between measurements.

RTI: Research Triangle Institute, a not-for-profit university-affiliated research organization in North Carolina that conducted the fifth round (the fourth follow-up in 2000) and the Postsecondary Education Transcript Study of NELS:88 for the National Center for Education Statistics. (See also entry for "PETS.")

Sample: Subgroup selected from the entire population.

Sampling error: The part of the difference between a value for an entire population and an estimate of that value derived from a probability sample that results from observing only a sample of values.

Sampling variance: A measure of dispersion of values of a statistic that would occur if the survey were repeated a large number of times using the same sample design, instrument and data collection methodology. The square root of the sampling variance is the standard error.

School administrator questionnaire: This questionnaire was to be completed by the principal and/or someone designated by the principal. The questionnaire sought basic information about school policies, number of students in each class, curriculum offered, programs for disadvantaged and disabled students, and other school characteristics.

School climate: The social system and culture of the school, including the organizational structure of the school and values and expectations within it.

School Coordinator: A person designated in each school to act as a contact person between the school and NORC. This person assisted with establishing a survey day in the school, and in some cases where the school cluster size was very small, the School Coordinator administered the student instruments.

Standard deviation: The most widely used measure of dispersion of a frequency distribution. It is equal to the positive square root of the population variance.

Standard error: The positive square root of the sampling variance. It is a measure of the dispersion of the sampling distribution of a statistic. Standard errors are used to establish confidence intervals for the statistics being analyzed.

Statistical significance: The finding (based on a derived probability, rather than an certitude) that two or more estimates are truly different from one, and not a merely apparent difference reflecting chance variation.

Stopout: A student who had one or more occurrences of school nonattendance for 20 or more days (not including any excused absences) who subsequently returned to school. In NELS:88, this term was used for temporary dropouts within a round (e.g., out of school in fall 1989 but back spring 1990, as contrasted to 1990 dropouts who were back in school in spring term of 1992).

Stratification: In a stratified sample, the total population is divided into strata or subgroups. Stratification is used to reduce sampling error. In NELS:88, the sampling frame was sorted to create strata or subgroups of schools and schools were selected independently within each stratum. Schools were stratified by superstrata (combinations of school type and geographic reason) and substrata (urban, suburban, rural; high versus low minority public schools).

Student questionnaire: One of the two parts of the student survey (the other part is the cognitive test battery). This instrument contained a locator section for tracing sample members for future waves of NELS:88 and a series of questions about courses taken, hours spent on homework, and perceptions of the school and the home environment.

Survey day: A day chosen by the school during the data collection period when an NORC interviewer and a clerical assistant (or the School Coordinator in schools with only a small group of sample members) administered the survey to the school's sample of students. The survey day session lasted about three hours for the actual data collection, with about thirty minutes each for preparation and clean-up/preparation of completed materials for mailing.

Teacher questionnaire: Math and science teachers of selected students were asked to complete a teacher questionnaire, which collected data on school and teacher characteristics (including teacher qualifications and experience), evaluations of student performance, and classroom teaching practices.

Teacher, NELS-targeted teacher sample: In the base year and first follow-up, two teacher reports were sought for each student, reflecting a combination of two subjects from four subject areas (English, social studies, science, mathematics). In the second follow-up, one teacher report per pupil was sought for those students who were enrolled mathematics, science, or both, in one of the schools designated for school contextual data collection.

Teacher transcript study: As a measure of the background and quality of teachers instructing NELS:88 8th-graders, postsecondary transcripts were collected for science and mathematics teachers of base-year students.

Tracing: The locating (and ascertaining of school enrollment status) of NELS:88 sample members. Sample members were traced at six points in time subsequent to 8th grade: autumn term 1988, autumn term 1989, spring term 1990, autumn term 1990, autumn term 1991, and spring term 1992.

Transfer student: A NELS:88 sample member who moved from one school to another after the subsampling of schools between phase 1 (the tracing of sample members to their school of enrollment) and phase 2 (the re-verification of sample members' school of enrollment).

Unit nonresponse: Failure of a survey unit (for example, at the institutional level, a school, or at the individual level, a respondent, such as a student or a teacher) to cooperate or complete survey instrument. Unit nonresponse may be contrasted to item nonresponse, which is the failure of a participating sample member to give a valid response to a particular question on a survey instrument.

Validity: The capacity of an item or measuring instrument to measure what it was designed to measure; stated most often in terms of the correlation between scores in the instrument and measures of performance on some external criterion. Reliability, on the other hand, refers to consistency of measurement over time. (See entry for "Reliability.")

Variance: See entries for "Population variance" and "Sampling variance."

Weighted estimates: Estimates from a sample survey in which the sample data are statistically weighted (multiplied) by factors reflecting the sample design. The weights (referred to as sampling weights) are typically equal to the reciprocals of the overall selection probabilities, multiplied by a nonresponse or poststratification adjustment. Thus, for example, the 1,035 completed school administrator questionnaires in the NELS:88 base year represent a population of 38,774 schools. Individual completed cases (that is, base-year school administrator questionnaires) may "represent" anywhere from a minimum of 1.5 schools to a maximum of 387.3 schools. To take another example, 12,111 base-year questionnaire respondents reported themselves to be male, and a slightly greater number (12,244) reported themselves to be female. When these cases are multiplied by the nonresponse-adjusted student weights to yield a weighted percent that reflects the national population of 8th-graders, the estimate for males is 50.1 percent of the 1988 8th-grade cohort while females are estimated to comprise 49.9 percent of the nation's 1988 8th-graders.

Appendix F:
NELS:88/2000 Derived Variables

Appendix F
NELS:88/2000 Derived Variables

Derived variables (or composite variables as they were termed in earlier NELS:88 follow-up studies) are constructed to enhance substantive analyses. Since research questions frequently require independent or control variables such as the type of postsecondary institutions attended, an individual's racial and ethnic status or gender, or the type of high school or postsecondary degree attained, over the course of this longitudinal study, a very large set of classification variables has been carefully constructed and added to the study's data files. In this appendix, we describe the derived variables developed for the fourth follow-up study from (a) the fourth follow-up study's interview, and (b) data collected during earlier waves of the study which were pulled from the NELS:88/94 restricted-use ECB, and NCES' 1997-98 Integrated Postsecondary Education Data System Institutional Characteristics file (IPEDS-IC). Our presentation includes a description for each of the derived variables; additional specification can be found in the reproducible SAS program used by the study team used to develop these variables.[95]

F.1 Derived Variables from Earlier Waves of NELS:88

Similar descriptions for the derived variables developed for the base-year and first three follow-up studies have been published and discussed elsewhere and are not repeated here. Interested readers can find these estimates in the following publications.

- Base-Year Study: A discussion of the composite variables for the NELS:88 base-year study can be found in the data file user's manual for the study's base-year student component (Ingels, Abraham, Karr, Spencer, and Frankel 1990, NCES 90-464). Section 7.2.4 in that report includes a description of these variables, and appendix D contains additional information on the variables.
- First Follow-up Study: Analysts interested in gaining more information on the composite variables used for the first follow-up study in 1992 should review the student component data file user's manual for this follow-up (Ingels, Scott, Lindmark, Frankel, and Myers 1992, NCES 92-030). In this report, section 7.2.4 and appendix H describe this wave's derived variables.
- Second Follow-up Study: For the second follow-up study conducted in 1992, analysts will find information on derived variables in this follow-up study's data file user's manual (Ingels, Dowd, Baldridge, Stipe, Bartot, and Frankel 1994, NCES 94-374). Section 7.2.3 describes these second follow-up variables; appendix H provides additional information.
- Third Follow-up Study: Chapter 7 of the NELS:88/94 methodology report (Haggerty, Dugoni, Reed, Cederlund, and Taylor 1996*)* includes information on this wave's composite/ derived variables.

F.2 Developing Derived Variables for the Fourth Follow-up Study

Most composite variables were constructed from two or more sources, and they may combine questionnaire items from the same or different NELS:88 data files, as well as from the same survey year or across different survey waves. Some composites are drawn from an external sampling resource that is

[95] An electronic version of the program (DerivedVarProg.sas) can be found in the "Derived Folder" on the CD-ROM containing the fourth follow-up study's electronic codebook.

unavailable to users, or use an external conceptual scheme in order to rank order or otherwise recode survey data. For example, whether a respondent ever attended a 4-year school required that the postsecondary institutions reported by the sample members be assigned an associated IPEDS unit ID; these IDs were then matched to the IPEDS-IC file to determine the school's educational level (e.g., less than 2-year or 4-year school).[96] A few composites are sufficiently central to analyses that they have been constructed in each round of the survey (e.g., high school graduation status). Some values should change over time; for example, if a sample member receives additional education or attains a higher degree.

Some variables, such as race/ethnicity and gender, should in theory be constant for an individual over time, yet in practice may change if new information updates the old. For example, regardless of actual participation in NELS:88, a race/ethnicity composite is constructed for all sample members. In a situation where a former nonparticipant later takes part in the survey, the value of the race composite may in very rare instances change from a value that had been imputed on earlier datasets. Such differences illustrate how the validity of certain classification variables is strengthened over time. In terms of these variables, the most recent round contains the best information for sample members who participated in that wave of NELS:88.

It should be noted, also with regard to the collection of information on race and ethnicity, that approaches to the collection of these data—and the definitions used to define the terms—have changed over time. This was certainly the case for the NELS:88 fourth follow-up study, which employed newly developed federal standards for collecting and reporting race and ethnic status.[97] To address the different needs of analysts, the NELS:88/2000 research team developed three derived variables for race/ethnicity. Two variables (F4RACEM and F4RACE2) employ the new guidance, but with slightly different conceptions of race. Another derived variable, F4RACE, is based on earlier conceptions for this term used in the second and third follow-up studies. We describe these variables and others in the following section.

F.3 Descriptions of Fourth Follow-up Study Derived Variables

SEX (F4SEX): Sex (male or female) of student is taken directly from the gender question in the NELS:88 second follow-up interview (F2SEX).

RACE/ETHNICITY (F4RACEM): This variable reflects new federal standards for collecting race and ethnicity data and allows the respondent to mark more than one choice for race. It is based on responses to F4JHISP ("Are you Hispanic or Latino") and to F4JRAC1 through F4JRAC3 ("Please select one or more of the following choices to best describe your race") in the NELS:88 fourth follow-up interview. If the respondents indicated that they were of Hispanic origin, they were assigned to the "Hispanic or Latino" category. If they responded that they were not Hispanic, their response to F4JRAC1 was taken as their primary choice of race/ethnicity. If they then had a valid value for F4JRAC2 or F4JRAC3, they were put in the "More than one race" category. Thus, the resulting categories of this variable are: 1) Asian or Pacific Islander; 2) Hispanic or Latino (any race); 3) Black, non-Hispanic; 4) White, non-Hispanic; 5) American Indian or Alaska Native; and 6) More than one race.

RACE/ETHNICITY (F4RACE2): This item reflects new federal standards for collecting race and ethnicity data. In contrast to F4RACEM, this composite variable is based solely on responses to F4JHISP

[96] Because of potential risk from these IPEDS unit identification numbers, these values appear only on the NELS:88/2000 restricted-use ECB.

[97] On October 30, 1997, the Office of Management and Budget (OMB) published new guidance for collecting and reporting data in federally supported research efforts. See, for example, "Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity" (*Federal Register*, 62 FR 58781–58790).

("Are you Hispanic or Latino?") and the first race variable (F4JRAC1– "Please select one or more of the following choices to best describe your race.") If they indicated they were of Hispanic origin their value on this variable was coded as 5 (Hispanic or Latino). If they responded that they were not of Hispanic origin, their response to F4JRAC1 was taken as their primary choice of race/ethnicity. Thus, the resulting categories of this variable are 1) American Indian or Alaska Native; 2) Asian or Pacific Islander; 3) Black, non-Hispanic; 4) White, non-Hispanic; and 5) Hispanic or Latino (any race).

RACE/ETHNICITY (F4RACE): This item describes sample member race and ethnic status by the approach employed during earlier rounds of the NELS:88 study. The variable is based on responses from each round of the study, with most data coming from the second follow-up study F2RACE1 variable. Unlike the other derived variables describing race-ethnicity, multiracial responses were not allowed.

HIGH SCHOOL COMPLETION STATUS BY 2000 (F4HSDIPL): This derived variable combines responses from the NELS:88 fourth follow-up interview and previous waves of the study to indicate sample members' current high school completion status. Sample member graduation status in 1994 (HSSTAT) was expanded with the fourth follow-up interview (F4DHSFIN) to identify sample members who had received diplomas or the equivalent. F4DHSACT provided the current educational activities for non-graduates in 2000. Three categories were generated for this variable: 1) Received high school diploma or equivalent; 2) Currently working towards a high school diploma/equivalent; and 3) Not received a diploma/equivalent and not working towards one.

HIGHEST POSTSECONDARY DEGREE ATTAINED BY 2000 (F4HHDG): This variable was derived from a set of questions that asked the respondents in the NELS:88 fourth follow-up interview about various types of postsecondary degrees or certificates they had attained by 2000. Four categories were generated for this variable: 1) Had no postsecondary experience; 2) Had some postsecondary education but not attained bachelor's degree; 3) Attained a bachelor's degree; and 4) Attained a master's degree or higher.

CURRENT EMPLOYMENT AND POSTSECONDARY ATTENDANCE STATUS (F4STATUS): The variable was derived from responses regarding current employment status (F4AACTF and F4AACTP) and postsecondary attendance status (F4AACTG) at the time of interview. Four categories were created: 1) Respondent was working for pay and not enrolled in school; 2) Respondent was enrolled in school and not working for pay; 3) Respondent was working for pay and enrolled in school; and 4) Respondent was neither working for pay nor enrolled in school.

SINGLE-PARENT STATUS (F4SGPAR): This variable was derived from respondent's marital status (F4GMRS) and whether the respondent lived with one or more children in the household (F4GNCHD) at the time of interview in the NELS:88 fourth follow-up interview. If a respondent reported that they were not married but living with one or more children or stepchildren, the respondent was considered as single parent.

TYPE OF HIGH SCHOOL DIPLOMA RECEIVED AS OF 2000 (F4HSTYPE): This variable was derived from a set of questions that asked NELS:88 fourth follow-up interview respondents about the type of high school diploma they had received. During the fourth follow-up interview, F4DHSFIN and F4DTYPE collected information on high school completion since the previous follow-up interview in 1994. High school degree attainment from the previous wave (HSSTAT) was collected from third follow-up study data. Three response categories were developed to describe degree type: 1) high school diploma, 2) GED certificate, and 3) certificate of attendance.

DATE OF RECEIPT OF HIGH SCHOOL DIPLOMA OR EQUIVALENT (F4HSGRDT): This variable was derived from questions that asked NELS:88 fourth follow-up interview respondents when they received their high school degrees. Respondents who provided this information during earlier data collection waves skipped this item, and their degree dates were taken from the NELS:88/94 data

(YRRECM and YRRECY). The fourth follow-up interview item F4DHSG provided the date for other respondents completing high school after the previous interview in 1994. (For those who went on to postsecondary institutions, another possible source of date [and type] of diploma is the postsecondary transcript. In some cases, information from these sources may disagree.)

TYPE OF POSTSECONDARY EDUCATION DEGREE ATTAINED AS OF 2000 (F4TYPEDG): This variable is derived from a series of questionnaire items that asked sample members to describe the degrees earned from all postsecondary institutions they had attended since high school. Information on degrees earned was based on responses from both the third follow-up interview (TYPDEGCT) and fourth follow-up interview (F4EDGR1–F4EDGR6). Students were able to select from up to six degree categories, including postsecondary certificate, associate's degree, bachelor's degree, master's degree, professional and academic doctorates. (Degree information is also reported on the postsecondary transcript file; because self-reports may differ from academic records, and because transcripts were generally collected at a date subsequent to the telephone interview, information from these sources may in some cases disagree.) From these degree types reported by the sample member during the third and fourth follow-up interviews, the derived variable included 10 categories, based on the level of degree reported. These categories included:

1. PSE experience, but no degree attained;
2. Attained certificate only;
3. Attained associate's degree only;
4. Attained bachelor's degree only;
5. Attained certificate and associate's degree, but not higher;
6. Attained certificate and bachelor's degree, but not higher;
7. Attained associate's and bachelor's degree, but not higher;
8. Attained certificate, associate's degree, and bachelor's degree, but not higher;
9. Attained master's degree but not higher; and
10. Attained academic or professional doctorate.

EVER ATTENDED A FOUR-YEAR POSTSECONDARY INSTITUTION (F4ATT4YR): This variable was derived from several measures describing the characteristics of the postsecondary education institutions that NELS:88/2000 respondents reported attending during either the third or fourth follow-up study interviews. Institutional level—the length of time required to complete the highest level of educational program offered by the school (i.e., less than 2 years, 2-3 years, 4 years and more)—was obtained from NCES' 1997-98 Integrated Postsecondary Education Data System Institutional Characteristics file (IPEDS-IC) for these schools. To reduce the potential for disclosure of individual respondents, institution-identifying information (e.g., UNITID for the IPEDS-IC file, F4LEVL, and F4CTRL) are included only on the NELS:88/2000 restricted-use ECB.

CURRENTLY ENROLLED IN A POSTSECONDARY INSTITUTION (F4ENRL00): This variable was derived from responses to the fourth follow-up study interview item (F4ELMY) that asked respondents to report the data that they "most recently attended postsecondary school." Sample members who were enrolled in postsecondary education during 2000 have values of 1.

F.4 Derived Variable Program

An electronic version of the program used to create the above derived variables is included on the CD-ROM containing the ECB. While this program is reproducible, the user should note that this program

draws from three external sources of data: (1) the public- or restricted-use version of the NELS:88/2000 ECB (N0P or N0R); (2) the restricted-use version of the ECB for the NELS:88 third follow-up study[98] (N4R), and (3) the 1997-98 IPEDS institutional characteristics file. Information on obtaining these datasets or restricted data licenses can be found at the NCES Web Site http://nces.ed.gov/.

[98] The public-use versions of the NELS:88/94 and NELS:88/2000 ECBs (N4P and N0P) do not include unique identification numbers (IPEDS unit IDs) for the postsecondary institutions that sample members reported attending. The public-use ECBs will be appropriate for duplicating the other derived variables in the program after the analyst correctly defines variable position and format—these elements differ between the two data files.