NATIONAL CENTER FOR EDUCATION STATISTICS

User's Manual September 1994

NATIONAL EDUCATION LONGITUDINAL STUDY OF 1988

SECOND FOLLOW-UP: STUDENT COMPONENT DATA FILE USER'S MANUAL

U.S. Department of Education
Office of Educational Research and Improvement NCES 94-374

NATIONAL CENTER FOR EDUCATION STATISTICS

User's Manual **September 1994**

NATIONAL EDUCATION LONGITUDINAL STUDY OF 1988

SECOND FOLLOW-UP: STUDENT COMPONENT DATA FILE USER'S MANUAL

Steven J. Ingels
Kathryn L. Dowd
John D. Baldridge
James L. Stipe
Virginia H. Bartot
Martin R. Frankel

National Opinion Research Center (**NORC**)
at the University of Chicago

Peggy Quinn
Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement **NCES 94-374**

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon Robinson
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Commissioner of Education Statistics

National Center for Education Statistics

"The purpose of the Center shall be to collect, and analyze, and disseminate statistics and other data related to education in the United Sates and in other nations."--Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e-1).

September 1994

Contact:
Peggy Quinn
(202) 219-1743

Foreword

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

Use of the data set does not require the analyst to be a sophisticated statistician or computer programmer. Most social scientists and policy analysts should find the data set organized and equipped in a manner that facilitates straightforward production of statistical summaries and analyses. This manual provides extensive documentation of the content of the data file and how to use it. **Chapter VII and Appendix I, in particular, contain essential information that allows the user to immediately proceed with minimal startup cost. A careful reading of Chapter VII and Appendix I will help users to avoid common mistakes that result in costly computer job failures or incorrect results.**

The rest of the manual provides a wide range of information on the design and conduct of the National Education Longitudinal Study of 1988 (NELS:88). Chapter I begins with an overview and history of NCES's National Education Longitudinal Studies program and the various studies that it comprises. Chapter II contains a general description of the data collection instruments used in the NELS:88 second follow-up.

The sample design and weighting procedures used in the second follow-up study are documented in Chapter III, as well as standard errors and design effects, non-sampling measurement errors, and problematic variables.

Data collection procedures, schedules, and results are presented in Chapter IV. Chapter V describes data control and preparation activities such as monitoring receipt of questionnaires, editing, and data retrieval. Chapter VI describes data processing activities including machine editing and construction of the cleaned data tape. Finally, Chapter VII describes the organization and contents of the data file and provides important suggestions for using it.

The appendices contain a list of other NCES NELS:88 publications; guidelines for Statistical Analysis System (SAS) users; the second follow-up student questionnaire; the record layout for the student questionnaire; specifications for the composite variables; the content areas of the second follow-up components; a glossary of project terms; a discussion of conducting cross-cohort trend analyses of students; and a codebook for the student questionnaire data.

In addition to the study described in this manual, a number of supplemental NELS:88 components are also described in Appendix A.

Earlier NCES longitudinal studies that may be of interest to NELS:88 users are described in Appendix B including the following: the High School and Beyond (HS&B) base year files; merged HS&B first, second, third, and fourth follow-up files; related HS&B files; and assorted files related to the National Longitudinal Study of the High School Class of 1972 (NLS-72).

A Note on Data Use and Confidentiality

The NELS:88 second follow-up data files are released in accordance with the provisions of the General Education Provisions Act (GEPA) [20-USC 122e 1] and the Carl D. Perkins Vocational Education Act. The GEPA assures privacy by ensuring that respondents will never be individually identified.

The National Center for Education Statistics (NCES) is responsible under the Privacy Act and Public Law 100-297 for protecting the confidentiality of individually identifiable respondents, and is releasing this data set to be used for statistical purposes only. Record matching or deductive disclosure by any user is prohibited.

To ensure that the confidentiality provisions contained in PL 100-297 and the Privacy Act have been fully implemented, procedures commonly applied for disclosure avoidance in other Government-sponsored surveys were used in preparing the data file associated with this manual. These include suppressing, abridging, and recoding identifiable variables. Every effort has been made to provide the maximum research information that is consistent with reasonable confidentiality protection. Deleted, abridged, and/or recoded variables appear with an explanatory footnote in the codebook attached to each user's manual.

Acknowledgements

A study such as this is built first and foremost upon the students, dropouts, teachers, school administrators, and parents who have so generously provided its basic data. We are grateful for their cooperation. We also thank the considerable numbers of school personnel who have assisted in the implementation of NELS:88.

We wish to acknowledge the role of a number of other individuals in the realization of the aims of this study. Donald Rock and Judith Pollack of Educational Testing Service served as task leaders for cognitive test development. Miriam Clarke provided counsel on management issues in the main study. Leslie Scott contributed significantly to the conceptualization and development of file specifications and composite variables for the components of the study.

We are also grateful to the members of NCES staff in the Longitudinal and Household Studies Branch who worked closely with us on this project. Jeffrey Owings, chief of the Longitudinal and Household Studies Branch; Peggy Quinn, project officer for the second follow-up; as well as other branch staff--Ralph Lee, Shi-Chang Wu, and Jerry West--who contributed to various aspects of this study. Bob Burton of the Statistical Standards and Methodology Division supplied statistical advice and review.

Three individuals in other agencies have worked particularly hard and effectively to help realize and extend the potential of NELS:88. Larry Suter of the National Science Foundation, Dick Berry (formerly of the National Science Foundation), and Carmen Simich-Dudgeon (formerly of the Office of Bilingual Education and Minority Languages Affairs (OBEMLA) of the U.S. Department of Education). We are grateful for their efforts.

In addition, we would like to express our appreciation to the members of what began in the base year as our National Advisory Panel, and became in 1989 the NELS:88 Technical Review Panel. The panelists--Jerald G. Bachman, Gordon Ensign, Lyle V. Jones, Nancy Karweit, Richard J. Murnane, Patricia Shell, Marshall S. Smith, and John Stiglmeier--provided wise counsel on many difficult issues of design, instrumentation and implementation. As consultants to the second follow-up, Aaron Pallas, Joan Talbert, Leigh Burstein, Anthony Bryk, and Senta Raizen also contributed importantly to the design and ultimate success of the study.

Steven J. Ingels was overall NELS:88 second follow-up project director. Lisa Thalji was associate project director responsible for securing school cooperation and locating NELS:88 cohort members. Katy Dowd was associate project director responsible for the student component during data collection. Laura Reed and Virginia Bartot were the data processing managers, and Martin R. Frankel was the task leader for sampling and statistics.

The authors also wish to acknowledge those who contributed to the production of this manual. Kenneth A. Rasinski performed the confidentiality disclosure analysis for the NELS:88 Second Follow-Up. Additionally, Doug Barge, Michael Ma, Gloria Rauens, Supriti Sehra, Shiow-Ling Tsai-Ma, and Hsiuling Young provided a great deal of their time and expertise to produce the statistics reported throughout the manual. Our appreciation is also extended to Karen Sutherlin and Cynthia Mathews for their patience and thoroughness in the production of the manuscript. Finally, we would like to thank the National Opinion Research Center field and telephone center interviewers and supervisors who with such energy and determination collected the NELS:88 data.

Table of Contents

Appendices

III. Sample Design and Implementation; Survey Error Assessment

This chapter describes the design and procedures used for selecting schools and students into the NELS:88 base year and first and second follow-up samples. It provides information on the calculation of sample weights and the relative efficiency of the sample design. The chapter also provides information about procedures used to adjust sample weights for nonresponse and about the effect of unit and item nonresponse and other potential sources of bias on estimates.

3.1 NELS:88 Sample Design

The following section describes the sample design of NELS:88, from its base year inception through the first and second follow-ups. Beginning from a straightforward two-stage stratified sample, the complexities of the NELS:88 sample design have grown exponentially with each subsequent wave.

3.1.1 Base Year Sample Design

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 eighth-grade enrollment to achieve virtual self-weighting. In addition, schools were oversampled in certain special strata so that policy-relevant subgroups would be adequately represented in the sample. Within each school approximately 26 students were to be randomly selected (typically, 24 regularly sampled students and two, on average, OBEMLA-supplement Hispanic and Asian/Pacific Islander oversampled students). In schools with fewer than 24 eighth 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 (or 23 participating students). From a national frame of about 39,000 schools with eighth grades, a target sample size of 1,032 schools was set. Some 1,052 schools--815 public and 237 private--participated and provided usable eighth-grade student data.

NORC's sampling frame was the school database compiled by Quality Education Data, Inc. (QED) of Denver, Colorado. The QED list contained information about whether a school was urban, suburban, or rural. NORC used this information for stratification purposes. The QED list did not at that time contain information about the racial/ethnic composition of individual public schools usable for the NELS:88 sampling frame. Racial/ethnic composition data were obtained from Westat, Inc. in its capacity as a NORC subcontractor for the NELS:88 base year study. As part of their work on the National Assessment of Educational Progress (NAEP), Westat had obtained data from the Office of Civil Rights (OCR) and from other sources (e.g., district personnel) that identified those schools with a minority enrollment of greater than 19 percent. Use of this data set facilitated the explicit stratification and allocation of schools with very large percentages of black or Hispanic students. Stratification information on whether a school was public, Catholic (private), or other private was obtained from the QED list and lists of private schools. Readers who desire more detail on the base year sample design should consult the *NELS:88 Base Year Sample Design Report*.

3.1.2 First Follow-Up Sample Design

There were three basic objectives for the NELS:88 first follow-up sample design. First, the sample was to include approximately 21,500 students who were in the eighth-grade sample in 1988 (including base year nonrespondents). This longitudinal cohort was to be distributed across 1,500 schools. Second, the sample was to constitute a valid probability sample of all students currently enrolled in the tenth grade in the 1989-1990 school year. This entailed freshening the sample with students who were tenth graders in 1990 but not in the eighth grade during the 1987-1988 school year. Third, the first follow-up was to include a sample of students who had been deemed ineligible for base year data collection (because physical, mental, or linguistic barriers prevented them from participating) so that those able to take part could be added to the first follow-up student sample, and demographic and school enrollment information could be obtained for them.

Longitudinal Cohort. The general sample design strategy for this component of the sample involved subsampling students selected for the base year with non-zero 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 equal to one). 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 one.

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 one).[1]

[1]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.

It was determined that the remaining pool of 25,988 students were distributed among 3,967 schools.[2] 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 with certainty. The remaining 6,420 students were distributed among 3,059 schools with 10 or fewer members of the base year sample. Their 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.[3]

Freshened Sophomore Sample. The second sampling objective was to create a valid probability sample of students enrolled in tenth grade in the 1989-1990 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 tenth-grade student selected for interview in 1990, a complete alphabetical roster of all tenth-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 eighth grade in the U.S. in 1988, then the freshening process terminated. If the designated student was not enrolled in the eighth grade in the U.S. 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

[2]When the school a student was attending could not be identified, a separate "school" of size one 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.

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

freshened sample) until a student who was in the eighth grade in the U.S. in 1988 was reached on the roster.

The freshening process could yield zero, one, or more than one new sample member in a given school. Altogether, 1,229 new students were added to the tenth-grade sample--on average, just less than one student per school. Some of these freshened students were dropped in the subsampling process (described below) either because they themselves 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.

Subsampling the Eighth-Grade Cohort and Freshened Sophomore Samples. After the initial selection of the longitudinal cohort, the combined longitudinal-freshened sample was further subsampled. 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 fifty 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 whom 357 were retained in the sample and pursued in the final data collection period of the study. In the course of final data collection, we did indeed find that 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 tenth-grade freshened student samples were reduced by 1,990 cases, yielding a first follow-up sample size of 20,706 (see Table 3.1.2-1).[4] 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. Of the revised 20,840 sample, 855 represent the first follow-up freshened sample, 19,645 represent the longitudinal cohort that began with eighth graders in 1988, 312 represent the base year ineligibles later found to be eligible, and 28 represent the base year ineligibles found to be out-of-scope.

Sample of Base Year Ineligibles. 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 are not proficient in English) and students whose physical or emotional problems would have made participation in the survey 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. Further detail on sample eligibility in the base year is provided in the *NELS:88 Base Year Sample Design Report*.

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 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

[4] The provisional first follow-up sample size of 20,706 has been 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 3.1.2-1
First follow-up sample by race breakdown[a]

	First Follow-Up Initial Selections	Freshened Sample	Dropped in final Subsampling[b]	Final Sample
All	21,474	1,229	1,997	20,706[c]
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
Whites	14,349	554	1,061	13,842
Missing/Refused	387	77	160	304

a Figures in this table 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 codebooks. This variable was used because it was the only race variable that was constructed for initial sample members dropped in final subsampling.

b 1,821 members of the eighth-grade longitudinal cohort and 169 freshened tenth graders were dropped in Phase 3 subsampling. In addition, 7 members of the eighth-grade longitudinal cohort were discarded because they were selected in error during the base year.

c This table is based on the original (1992-1993) 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. Additional details about the sample numbers of the two releases are on page 31 of section 3.1.2, under the subheading "Subsampling the Eighth-Grade Cohort and Freshened Sophomore Samples."

eligibility change, as another means of providing for a probability sample of 1992 twelfth graders.[5] Students whose ineligibility

[5] 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 eighth graders had become ineligible in the first or second follow-ups (for example, because of mental or physical problems engendered by an accident). We have 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 retain their overall

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, collecting additional demographic information would help to better describe any undercoverage biases, while collecting school enrollment status information would facilitate a more accurate estimation of a national dropout rate between grades eight and ten.

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, we applied the same within-school sampling rates that had been used in selecting the base year sample students. 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 electronic codebook. Student questionnaire data from those who were successfully surveyed are included in the combined base year/first follow-up/second follow-up data release. For details of the sampling methodology and composition of the base year ineligibles sample, see the *NELS:88 First Follow-Up Final Technical Report*; for a statement of the data analysis implications of undercoverage of the limited English language proficient population, see section 3.4.1 of this manual.

3.1.3 Second Follow-Up Sample Design

There were five basic objectives for the NELS:88 second follow-up sample design. **First**, the sample was to constitute a valid probability sample of all students enrolled in the twelfth grade in the 1991-1992 school year. This entailed freshening the sample with students who were twelfth graders in 1992 but were not in the eighth grade in the U.S. in the 1987-88 school year, just as the first follow-up sample had been freshened in 1989 to achieve a 1990-91 representative sample of sophomores. Additionally, it was necessary to reassess the eligibility status of selected students found in previous waves to be ineligible, and to include them in the cohort if they were determined to be eligible for the second follow-up. **Second**, to continue the examination of the dropping out phenomenon, dropouts were to be retained with certainty. **Third**, it was highly desirable for policy analysis purposes to retain the maximum number of Hispanics, Asians, and American Indians from the first follow-up sample. **Fourth**, to minimize nonresponse bias first

sample eligibility. Future waves of NELS:88 may wish to reassess their eligibility for participation in those data collection efforts.

follow-up nonrespondents were to be retained with certainty. **Fifth**, the sample was to be clustered in 1,500 schools from which contextual data--including 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.

Longitudinal Cohort. 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. After eliminating the locations of the "known" dropouts[6] (*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.[7]

It was clear that even if no attempt were made to satisfy the second goal--retention with near certainty of Hispanics, Asians, and American Indians from the first follow-up sample--that the fifth goal of achieving a cluster of students in 1,500 schools could not be met without significant losses in sample efficiency,

---

[6] 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 1988-1990 dropout episode), regardless of their school enrollment status as of the second follow-up spring term 1991 tracing effort. In other words, dropouts who had since returned to school and stopouts who remained in school were still counted as dropouts for sampling purposes, along with institutionalized individuals and the additional dropouts identified during second follow-up tracing. Some dropouts for sampling purposes who were out of school after tracing returned to school and were interviewed as spring term 1992 students.

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

effective sample size, or both. Table 3.1.3-1 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. The data in the table indicated that 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 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.

Teacher, school administrator, and transcript components were limited to a maximum of 1,500 schools. For this reason it was still necessary to select a sample of schools, although the students falling outside that sample would not be excluded from the study. For students in the 1,500 schools selected, the full range of data--student, parent, teacher, school administrator, and transcript data--were collected; for the students in a school not among those selected, only student and parent data were collected.

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. Figure 3-1 provides an illustration of the longitudinal sample design of the base year and first follow-up, as well as that of the second follow-up.

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

Table 3.1.3-1
Clustering of first follow-up sample members eligible for second follow-up
(schools [*N*=2,258] and non-school locations)

School Size	Total Schools	Total Schools With API,HIS,AI	Total Schools Without
1	1974	579	1395
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

Table 3.1.3-1 (cont.)
Clustering of first follow-up sample members eligible for second follow-up
(schools [*N*=2,258] and non-school locations)

School Size	Total Schools	Total Schools With API,HIS,AI	Total Schools Without
45	1	1	0
50	1	1	0
53	1	1	0
60	1	1	0
Total	3224	1383	1841

Note: known school-leavers are not included in the numbers above.

attendance at schools with fewer NELS:88 students. For example, students who are more likely to transfer to different schools will be under-represented if the weight is not applied. Further details can be found in section 3.2 on second follow-up weighting.

Freshened Senior Sample. The sample freshening process was once again employed in the second follow-up to ensure that 1992 twelfth 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 1991-1992 high school seniors. (The second obstacle was the prior exclusion of some 1988 eighth graders, which is addressed in the next section.) The procedure was implemented in four steps as described in section 3.1.2 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 twelfth grade in 1992 but were not enrolled in the eighth grade in the U.S. in 1988. There is a very small bias introduced by the omission of eligible twelfth graders attending schools that

Figure 3-1: NELS:88 8th Grade Spring Defined Cohort Status Distribution in First and Second Follow-Ups

Base Year	First Follow-Up Status	Second Follow-Up Status	
Students N = 20,062	Dropouts N = 1,029	Dropout	N = 611
		Alt. Completer[a]	N = 222
		Student	N = 69
		Out of Scope	N = 9
		Status Unknown	N = 118
	Students N = 18,270	Dropout	N = 1,041
		Alt. Completer[a]	N = 542
		Student	N = 16,339
		Out of Scope	N = 82
		Status Unknown	N = 266
	Out of Scope N = 129	Dropout	N = 11
		Alt. Completer[a]	N = 6
		Student	N = 11
		Out of Scope	N = 83
		Status Unknown	N = 18
	Status Unknown N = 634	Dropout	N = 58
		Alt. Completer[a]	N = 20
		Student	N = 466
		Out of Scope	N = 6
		Status Unknown	N = 84

[a]Alt. Completer = Alternative Completer or Alternative Student
Note: In addition to the 20,062 sample members listed above, an additional 1,126 sample members were added due to sample freshening. Thus, 20,062 and 1,126 equals the 21,188 cases found on the data file tape.

included *no* students who were eighth graders in 1988.[8] There is an additional small bias introduced by not freshening on the members of the sample of base year ineligibles. All other 1992 twelfth graders who qualify for the freshening sample had some chance of selection. Because each 1988 eighth grader added through first follow-up freshening had a calculable, non-zero probability of selection into the base year sample, we can calculate the selection probabilities for all students eligible for the freshening sample. Thus, the freshening procedure produces a sample that meets the criterion 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 followback study of excluded students, this liberalized eligibility criterion was also applied to excluded 1987-88 eighth graders at two points in time.) Of the 366 students initially sampled through the freshened 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%) 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**.[9]

---

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

[9] 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 first and second follow-up. While these "freshened dropouts" were included on the original first follow-up public release, in the current re-release these

The school sample, for purposes of collecting contextual data, comprised the 1,387 schools that represent selected clusters (as traced in Phase 1) at which 1) NELS:88 sample members were still present in the 1991-92 school year, and 2) provided at least one completed student questionnaire.

Followback 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 followback study of excluded students. Two additional groups of students, however, 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. Second, a quite 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 eligible in a previous round(s) 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.

The second follow-up followback study of excluded students pursued essentially the same objectives as informed the first follow-up base year ineligible study. Since 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 counsellor, or English-as-a-Second-Language teacher). Additionally, complete school enrollment status information was obtained, as well as confirmation of basic demographic characteristics.

This approach implemented in the first and second follow-ups allows for some deviance from the traditional definition of survey participation and a special weight creation to calculate dropout rates adjusted for 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 is no completed questionnaire in a round, receive 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 presence in the study, based on whether school enrollment status information and the most basic sociodemographic classification variables can be obtained. Particularly for the generation of school retention and dropout statistics, and in order to statistically accommodate students who are incapable of participation in the most strict sense of questionnaire and test

---

cases appear only on the restricted use files.

completion (and those who are capable but did not participate) basic sociodemographic and school persistence information has been collected through school personnel or by proxy (usually a parent or guardian) for both nonparticipants and ineligibles. A special weight has been created to reflect this expanded definition of the "participating" population and can be applied to calculate, for example, adjusted national dropout rates for the periods between eighth, tenth and twelfth grades.

3.2 Calculation of Weights

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 is the process that was applied to weighting NELS:88 data in all rounds.

3.2.1 Calculation of Base Year Sample Weights

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 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 eighth graders in public, Catholic, and other private schools in 1988.

The base year weighting procedures consisted of two basic stages:

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

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

The nonresponse-adjusted school weight was derived as the product of the school's preliminary weight times a nonresponse adjustment factor intended to adjust for the fact that some of the sampled schools did not return a completed questionnaire. The preliminary weight for students was based upon the inverse of the probability that the student's school was selected into the sample multiplied by the inverse of the probability that the student was sampled within the school. The nonresponse-adjusted student weight was derived as the product of the student's preliminary weight times a nonresponse adjustment factor intended to adjust for the fact that some of the sampled students did not participate, that is, did not return a completed questionnaire. Statistical properties of the base year weights are presented in Table 3.2.1-1.

Each school appearing on the NELS:88 base year school file, and each student appearing on the NELS:88 student file, has a value for the final weight variable. The weight represents the probability of selection into the sample, in addition to a factor that adjusts for nonresponse. Thus, the weight serves the purpose of allowing a particular case to represent other nonsampled cases within its sampling stratum, and to represent nonresponding cases similar to it in various respects. Because separate final student and school weights have been provided, the construction of each will be considered separately in the following discussion.

Base Year School Weights. The final school weight, BYADMWT, was derived using a multistage process. First, an initial weight--which represented the inverse of the school's selection probability--was attached to each school record in a file containing records for all eligible schools in the NELS:88 sample. A logistic regression procedure was used to estimate (in terms of a probability of nonresponding) the degree to which each of the responding schools resembled a nonresponding school. This estimated probability of nonresponse was the first adjustment factor applied to a school's weight.

Next, a polishing procedure--multi-dimensional raking--further adjusted the weights to sum to known population totals within strata. Estimating the nonresponse probability for each of the responding schools was possible because key background information on almost all of the nonresponding schools was available.

The final result of these procedures was a weight for each of the responding schools adjusted to compensate for nonresponse. For the purpose of adjusting the school weight, a nonresponding school was defined as a school for which both school administrator questionnaire data and student questionnaire data were unavailable.

Base Year Student Weights. The final student weight, BYQWT, was also derived using a multistage process. A design weight for each eligible student on a participating school's sample roster represented the student's probability of selection within the school. A student-level nonresponse adjustment factor was calculated by forming weighting cells based upon the combination of certain levels of variables representing school type, region, ethnicity, and gender. For each student, the product of a preliminary school weight and the student's design weight was formed. (The preliminary school weight was slightly different from

Table 3.2.1-1
NELS:88 base year statistical properties of sample case weights

Weight	School BYADMWT	Student BYQWT
Mean	37.46	122.29
Variance	2,109.17	4,359.16
Standard deviation	45.92	66.02
Coefficient of variation (×100)	122.59	53.99
Minimum	1.54	2.44
Maximum	387.30	836.91
Skewness	2.69	2.18
Kurtosis	9.47	16.32
Sum	38,774.12	3,007,779
Number of cases	1,035	24,599

BYADMWT. BYADMWT was adjusted to accommodate the 17 schools for which school administrator questionnaire data were unavailable though student questionnaire data had been obtained. The preliminary school weight eliminated this step in the adjustment process. Thus, it is appropriate for application to the 1,052 schools with student questionnaire data available.) This product was summed for all students and all participating students within weighting cells. The ratio of the sums for all sampled students to participating students was used as the nonresponse adjustment factor for each student's design weight.

3.2.2 Calculation of 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 eighth grade during the 1987-88 school year or in the tenth grade during the 1989-90 school year. Thus, this population encompasses both populations of prime analytic interest--the population of 1990 tenth graders (including those who were not eighth graders in 1988) and the 1988 eighth-grade population (excluding any additional 1990 tenth 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 tenth 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 eighth 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 eighth graders deemed eligible for the base year survey, 1990 tenth graders who were not in the eighth grade in 1988, and 1988 eighth graders who were deemed ineligible for participation in the base year but were considered eligible to participate in the first follow-up.

Eligible 1988 Eighth 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 that reflected the selection probabilities for each case. Each case 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 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:

$$FFUDW_i = BYDW_i \times (1/P_{1i}),$$

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:

a. Out of sequence students (i.e., those who were not in tenth grade in 1990);
b. Dropouts identified at the time of initial first follow-up sampling;
c. Students who had transferred out of the first follow-up school from which they were selected;
d. Potential dropouts;
e. Other students initially classified as attending schools with 3 or fewer base year students; and,
f. 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, NAIS private, and other private), student sex (male and female), race/ethnicity (non-Hispanic white, American Indian, Hispanic, Asian, non-Hispanic black, and unknown), and base year 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 multidimensional raking.[10]

1990 Tenth Graders Who Were Not 1988 Eighth Graders. All members of this population who are included in the first follow-up sample were selected through the freshening process. This process linked each 1990 tenth grader who was not a 1988 eighth grader to a student who was an eighth 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 distributions for all cases selected through freshening (including nonrespondents). The two dimensions used in the raking procedure

[10] 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.

were sex and race/ethnicity (non-Hispanic white, American Indian, Hispanic, Asian, non-Hispanic black, and unknown as the categories).

1988 Ineligible Eighth 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 several of the second follow-up procedures. These procedures are discussed in more detail in section 3.2.3.

The first follow-up design weight was obtained by dividing the base year design weight by .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 section 3.2.3), gender and race (API/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 their cell. This adjusted weight serves as their F1QWT.

First Follow-Up Panel Weight (F1PNLWT). The panel weight was developed only for those cases who were selected for both the base year and first follow-up samples and who provided complete data in both rounds. The same procedures used in developing the basic first follow-up weight for 1988 eighth 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 cases who remained eligible for the first follow-up sample. The same six nonresponse adjustment groups and multidimensional raking procedures used in calculating the basic first follow-up weight were also used in calculating the panel weight.

Results of Weighting. To check the sample case weights, we analyzed the statistical properties of the weights; Table 3.2.2-1 displays the mean, variance, standard deviation, coefficient of variation, minimum, maximum, skewness, and kurtosis for both of the weights included on first follow-up data files.

Users should note that compared to the base year questionnaire weight (BYQWT), the first follow-up questionnaire (F1QWT) and panel (F1PNLWT) weights are larger, on average, and more variable. (For BYQWT, refer to Table 3.2.1-1.) This mostly reflects the effect of subsampling students at different rates depending upon the number of other NELS:88 students with whom they were clustered in their first follow-up schools.

Table 3.2.2-1
NELS:88 first follow-up statistical properties of sample weights for dropouts on the 1990 release of the first follow-up student files[a]

WEIGHT	F1QWT	F1PNLWT
Mean	165.88	172.62
Variance	46,249.54	52,603.86
Standard Deviation	215.06	229.36
Coefficient of Variation (×100)	129.65	132.86
Minimum	2.14	2.26
Maximum	6,996.81	7,479.71
Skewness	10.89	11.22
Kurtosis	205.24	214.14
Sum	3,217,069.00	3,007,813.00
Number of Cases	19,394.00	17,424.00

[a] This table is based on the original (1992-1993) 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. Additional details about the sample numbers of the two releases are on page 31 of section 3.1.2, under the subheading "Subsampling the Eighth-Grade Cohort and Freshened Sophomore Samples."

3.2.3 Calculation of Second Follow-Up Weights

Explanation of Weights. Eight weights were developed for inclusion on the 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 eighth grade during the 1987-88 school year or in the tenth grade during the 1989-90 school year, or in the twelfth 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 twelfth graders.

F2PNLWT This panel weight applies to sample members who completed a questionnaire in 1988, 1990, and 1992 (all three rounds of NELS:88). This can be used to make projections to the population of 1988 eighth 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 eighth grade in 1988 or in the tenth grade in 1990. By selecting appropriate sample members with the flag F2F1PNFL, analysts can use F2F1PNWT to make projections to such populations as 1990 tenth graders.

F2CXTWT This cross-sectional weight applies to students who attended the schools selected for inclusion in the 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 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 twelfth grade during the 1991-92 school year for whom we 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-up) and for whom transcript data are available. F2TRP2WT allows analysts to perform panel analyses using transcript data in conjunction with 1990-1992 test and questionnaire data.

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

Process for Calculation of Second Follow-Up Weights. A basic four-step process was defined for the calculation of all eight questionnaire weights. The first step, developing a classification scheme, was done at the beginning of the weighting process for all students in the sample. 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 High School and Beyond 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 non-traditional 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 non-traditional 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 due to a language barrier or a mental or physical incapacity.

"Expected grade" means tenth grade in the first follow-up and twelfth 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 twelfth graders in F2TRSCWT, the design weight used is equal to the first follow-up design weight.[11] 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 due to their association with a sampled school in F2TRSCWT and for all members in the F2CXTWT population, it 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 upon combinations of the classification values from step 1 as well as race (Hispanic, API, 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 their nonresponse adjusted weight. This step was performed independently for each weight calculated. For second follow-up freshened students the nonresponse adjusted weight serves as their final weight.

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:

a) Targets were developed that used the second follow-up

[11] 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.

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.[12] It was used in developing total sum of weights 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).

b) 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, API, American Indian, unknown), gender, base year school region, base year school type, and base year school urbanicity.

Results of Weighting. To check the sample case weights, the statistical properties of the weights were analyzed; Table 3.2.3-1 displays the mean, variance, standard deviation, coefficient of

[12] 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 (API, 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 the freshened). Second follow-up freshened students have their second follow-up design weight as their expanded sample weight. This step was performed for the sample as a whole.

variation, minimum, maximum, skewness, and kurtosis for the weights included on second follow-up student data files. Tables showing results for the remaining five weights can be found in the school (contextual weight), transcript (transcript weights), and parent (parent weight) data file user's manuals and the NELS:88 Second Follow-Up Sample Design Report.

Table 3.2.3-1
NELS:88 second follow-up statistical properties of sample weights for all sample members on student component public use data file

WEIGHT	F2QWT	F2PNLWT	F2F1PNWT
Mean	167.75	180.17	174.66
Variance	43,671.80	50,610.95	46,174.76
Standard Deviation	208.98	224.97	214.88
Coefficient of Variation (×100)	124.58	124.86	123.03
Minimum	2.14	2.39	2.31
Maximum	6,670.09	7,388.13	6,780.07
Skewness	10.18	11.59	10.63
Kurtosis	180.09	233.60	196.94
Sum	3,224,099	2,970,835	3,164,096
Number of Cases	19,220	16,489	18,116

3.3 Standard Errors and Design Effects

In this section we discuss the calculation of standard errors as a measure of sampling variability in survey results; the standard error is an estimate of the expected difference between a statistic from a particular sample and the corresponding population value.

Survey Standard Errors. Because the NELS:88 sample design involved stratification, disproportionate sampling of certain strata, and clustered (i.e. multi-stage) probability sampling, the resulting statistics 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 SPSS (Statistical Program for the Social Sciences) or SAS (Statistical Analysis System) do not calculate standard errors by taking into account complex sample designs. Several procedures are available for calculating precise estimates of sampling errors for complex samples. Procedures such as Taylor Series approximations, Balanced Repeated Replication (BRR), and Jackknife Repeated Replication (JRR) produce similar results.[13] Consequently, it is largely a matter of convenience which approach is taken. For NELS:88, NORC used the Taylor Series procedure to calculate the standard errors.

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. 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 design effects and root design effect for this section:

$$\text{DEFF} = \frac{(\text{DESIGN-SE})^2}{(\text{SRS-SE})^2} \qquad (1)$$

$$\text{DEFT} = \frac{\text{DESIGN-SE}}{\text{SRS-SE}} \qquad (2)$$

where DESIGN-SE designates the standard error of an estimate calculated by taking into account the complex nature of the survey

[13] Frankel, M.R., *Inference from Survey Samples: An Empirical Investigation* (Ann Arbor: Institute for Social Research, 1971).

design, and SRS-SE designates the standard error of the same estimate calculated as if the survey design was a simple random sample.

3.3.1 Base Year 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.[14] The 30 variables from the student questionnaire were selected to overlap as much as possible with those variables examined in High School and Beyond. 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 and ethnicity, school type (public, Catholic, and other private), and socioeconomic status (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 eighth-grade enrollment (at or below the median and above the median).

Results. Design effects for questions selected from the student questionnaire are presented in Table 3.3.1-1. On the whole, the design effects indicate that the NELS:88 sample was slightly more efficient than the High School and Beyond sample. For means and proportions based on student questionnaire data for all students (see Table 3.3.1-1), the average design effect in the NELS:88 base year was 2.54; the comparable base year figure was 2.88 for the High School and Beyond sophomore cohort and 2.69 for the senior cohort. Table 3.3.1-2 gives the mean design effects (DEFFs) and mean root design effects (DEFTs) for each subgroup. This table shows that the difference is also apparent for subgroup estimates. The *High School and Beyond Sample Design Report* presents design effects for ten subgroups defined similarly to those in Table 3.3.1-2.[15] For eight of the ten subgroups, the NELS:88 design effects are smaller on the average than those for both the High School and Beyond sophomore and senior cohorts. The increased efficiency is especially marked for students attending

[14] For a more detailed presentation of design effects for individual items for the total sample and for various subsamples, see the *NELS:88 Base Year Sample Design Report*. For tables of base year parent and school administrator questionnaire data standard errors and design effects, see the respective base year data file user's manuals, or the sample design report.

[15] Frankel, M; Kohnke, L.; Buonanno, D.; and Tourangeau, R.; Chicago: NORC, 1981.

Table 3.3.1-1
NELS:88 base year student questionnaire data: standard errors and design effects (*N*=24,599)

Survey item (or composite variable)	All Students	Estimate	Design S.E.[a]	DEFF	DEFT	N	SRS S.E.[b]
Mother/female guardian living	BYS2A	99.35	0.06	1.35	1.16	24126	0.05
Father/male guardian living	BYS7A	91.48	0.26	1.94	1.39	22775	0.19
Expect to attend public high school	BYS14	88.13	0.43	4.21	2.05	24156	0.21
Father finished college	BYS34A	29.36	0.65	4.18	2.04	20450	0.32
Mother finished college	BYS34B	22.94	0.50	3.03	1.74	21504	0.29
Parents require chores to be done	BYS38B	90.11	0.23	1.39	1.18	24392	0.19
Watch more than 2 hrs of TV per weekday	BYS42A	66.35	0.47	2.18	1.48	22042	0.32
I feel good about myself	BYS44A	92.26	0.23	1.73	1.31	24355	0.17
Good luck more important than hard work	BYS44C	11.87	0.25	1.48	1.22	24245	0.21
Every time I get ahead something stops me	BYS44F	28.50	0.40	1.87	1.37	24266	0.29
Plans hardly work out, makes me unhappy	BYS44G	20.16	0.34	1.78	1.34	24258	0.26
I feel I do not have much to be proud of	BYS44L	14.26	0.29	1.64	1.28	24200	0.22
Expects to finish college	BYS45	65.44	0.49	2.62	1.62	24384	0.30
Expects to graduate from high school	BYS46	98.20	0.10	1.46	1.21	24332	0.09
Talk to father about planning H.S. prgrms	BYS50A	73.98	0.41	2.05	1.43	23795	0.28
Student cutting class a problem at school	BYS58C	14.96	0.37	2.51	1.58	23849	0.23
Student use of alcohol a problem at school	BYS58G	15.32	0.35	2.23	1.49	23838	0.23
Parents wanted R to take algebra	BYS62	57.42	0.60	2.25	1.50	15084	0.40
Enrolled in advanced mathematics	BYS66D	41.09	0.51	2.46	1.57	23159	0.32
English will be useful in my future	BYS70C	84.14	0.30	1.60	1.26	23379	0.24
Afraid to ask questions in social studies	BYS71B	15.09	0.32	1.82	1.35	23225	0.23
Ever held back a grade in school	BYS74	17.66	0.37	2.12	1.46	22771	0.25
Often come to class without homework	BYS78C	21.86	0.34	1.60	1.26	23062	0.27
Participated in school varsity sports	BYS82B	47.85	0.57	2.96	1.72	22578	0.33
Participated in dance	BYS82G	26.67	0.50	2.86	1.69	22383	0.30
Participated in religious organization	BYS82T	14.89	0.34	2.07	1.44	22120	0.24
Reading test formula score	BYTXRFS[c]	10.23	0.08	4.12	2.03	23791	0.04
Mathematics test formula score	BYTXMFS[c]	15.98	0.16	4.99	2.23	23778	0.07
Science test formula score	BYTXSFS[c]	09.86	0.08	4.82	2.20	23765	0.04
History/government test formula score	BYTXHFS[c]	15.12	0.11	5.01	2.24	23673	0.05

Mean	2.54	1.56
Minimum	1.35	1.16
Maximum	5.01	2.24
Standard deviation	1.11	0.33
Median	2.15	1.47

[a] Standard error calculated taking into account the sample design.
[b] Standard error calculated under assumptions of random sampling.
[c] Although this table does not reflect the rescaling of base year cognitive test items in the second follow-up, the correlation between the cognitive test items before and after the rescaling is 0.99.

Table 3.3.1-2
Mean design effects (DEFFs) and root design effects (DEFTs) for base year student questionnaire data

Group	Mean DEFF	Mean DEFT
All students	2.54	1.56
Male[a]	1.98	1.39
Female	1.93	1.38
White and other[b]	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

[a]Sex categories are based on the composite sex variable.
[b]Race categories are based on the composite race variable.

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

Table 3.3.1-3 NELS:88 first follow-up:
Standard errors and design effects, all respondents; full sample (*N*=19,264)[a]

All Students and Dropouts

Survey item (or composite variable)		Estimate	Design S.E.[b]	DEFF	DEFT	N	SRS S.E.[c]
Sure to graduate from H.S	F1S18A	95.51	0.403	7.182	2.680	18945	0.150
Sts in collg Prep/acadmc pgm	F1S20C	31.56	0.784	5.362	2.315	18843	0.339
Sts in vocational/tec pgms	F1S20D	11.50	0.435	3.504	1.872	18843	0.232
Watch more than 2hrs/per weekdy	F1S45A	54.52	0.693	3.491	1.868	18026	0.371
Expect to finish college	F1S49	54.95	0.776	4.627	2.151	19023	0.361
At age 30 exp to be a manager	F1S53F	5.23	0.252	2.300	1.517	17959	0.166
At age 30 exp to be in the military	F1S53G	2.97	0.188	2.204	1.485	17959	0.127
At age 30 exp to be an operative	F1S53H	1.43	0.223	6.318	2.513	17959	0.089
At age 30 exp to be a clergyman	F1S53J	18.11	0.535	3.465	1.861	17959	0.287
At age 30 exp to be a technician	F1S53P	4.67	0.223	2.007	1.417	17959	0.157
At age 30 doesn't know what to be	F1S53S	10.47	0.365	5.376	2.319	17959	0.157
Others in home speak Spanish	F1S55	57.69	2.296	8.462	2.909	3919	0.789
I feel good about myself	F1S62A	91.99	0.292	2.083	1.443	18007	0.202
Luck is more imprtnt than hrd wk	F1S62C	12.64	0.460	3.427	1.851	17887	0.248
Something always prevnts success	F1S62F	27.90	0.607	3.277	1.810	17889	0.335
My plans do not work out	F1S62G	22.55	0.545	3.034	1.742	17837	0.313
I do not have much to be proud of	F1S62L	17.41	0.471	2.746	1.657	17800	0.284
Live with other adult male in hh	F1S92C	7.04	0.376	4.129	2.032	19109	0.185
Live with mother in same hh	F1S92D	88.39	0.463	3.991	1.998	19109	0.232
Live with stepmother in same hh	F1S92E	3.04	0.192	2.391	1.546	19109	0.124
Live with boy/girl friend	F1S92H	1.34	0.129	2.396	1.548	19109	0.083
Live with own children	F1S92I	3.69	0.235	2.970	1.723	19109	0.136
Parents require chores to be done	F1S100E	94.29	0.269	2.327	1.525	17324	0.176
#-Grandparents in same household	F1S93C	0.10	0.005	2.462	1.569	16672	0.003
#-Relatives under 18 in same hh	F1S93D	0.09	0.006	2.423	1.557	16625	0.004
#-Nonrelatives under 18 in hh	F1S93F	0.04	0.004	2.202	1.484	16578	0.003
Reading test formula score	F1TXRIR[d]	21.08	0.133	5.215	2.284	17832	0.058
Mathmtcs test formula score	F1TXMIR[d]	35.53	0.220	5.661	2.379	17793	0.092
Science test formula score	F1TXSIR[d]	13.68	0.090	5.581	2.362	17684	0.038
Hist/Cit/Geog test formula score	F1TXHIR[d]	18.94	0.098	5.121	2.263	17591	0.043

Mean	3.858	1.923
Minimum	2.007	1.417
Maximum	8.462	2.909
Standard deviation	1.681	0.408
Median	3.446	1.856

[a] This table is based on the original (1992-1993) 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 page 31 of section 3.1.2 for additional details about the sample numbers of the two releases.
[b] Standard error calculated taking into account the sample design.
[c] Standard error calculated under assumptions of simple random sampling.
[d] Although this table does not reflect the rescaling of first follow-up cognitive test items in the second follow-up, the correlation between the cognitive test items before and after the rescaling is 0.99.

**Table 3.3.1-4 NELS:88 first follow-up:
Standard errors and design effects, all respondents, panel sample (*N*=17,424)[a]**

All Students and Dropouts

Survey item (or composite variable)		Estimate	Design S.E.[b]	DEFF	DEFT	N	SRS S.E.[c]
Sure to graduate from H.S.	F1S18A	95.82	0.420	7.580	2.753	17208	0.153
STS in college prep/academic pgms	F1S20C	32.61	0.837	5.439	2.332	17065	0.359
STS in vocational/technical pgms	F1S20D	11.08	0.439	3.337	1.827	17065	0.240
Watch TV more than 2 hrs/per wkday	F1S45A	54.44	0.719	3.428	1.851	16448	0.388
Expect to finish college	F1S49	56.47	0.799	4.473	2.115	17223	0.378
At age 30 expect to be a manager	F1S53F	5.22	0.272	2.440	1.562	16333	0.174
At age 30 exp to be in the military	F1S53G	2.94	0.196	2.197	1.482	16333	0.132
At age 30 exp to be an operative	F1S53H	1.47	0.244	6.723	2.593	16333	0.094
At age 30 exp to be a clergyman	F1S53J	18.58	0.561	3.398	1.843	16333	0.304
At age 30 expect to be technician	F1S53P	4.63	0.215	1.708	1.307	16333	0.165
At age 30 doesn't know what to be	F1S53S	10.11	0.370	5.059	2.249	16333	0.165
Others in home speak Spanish	F1S55	57.59	2.232	6.921	2.631	3394	0.848
I feel good about myself	F1S62A	92.09	0.311	2.185	1.478	16450	0.210
Luck is more imp than hard work	F1S62C	12.12	0.458	3.218	1.794	16345	0.255
Something always prevents success	F1S62F	27.24	0.639	3.369	1.835	16351	0.348
My plans do not work out	F1S62G	21.92	0.557	2.955	1.719	16301	0.324
I do not have much to be proud of	F1S62L	16.79	0.471	2.583	1.607	16269	0.293
Live with other adult male in hh	F1S92C	6.85	0.410	4.558	2.135	17302	0.192
Live with mother in same hh	F1S92D	88.59	0.501	4.297	2.073	17302	0.242
Live with stepmother in same hh	F1S92E	3.11	0.213	2.607	1.615	17302	0.132
Live with boy/girl friend	F1S92H	1.28	0.136	2.527	1.589	17302	0.085
Live with own children	F1S92I	3.61	0.248	3.059	1.749	17302	0.142
Parents require chores to be done	F1S100E	94.52	0.277	2.350	1.533	15857	0.181
#-Grandparents in same household	F1S93C	0.10	0.005	2.390	1.546	15305	0.003
#-Relatives under 18 in same house	F1S93D	0.08	0.006	2.565	1.601	15264	0.004
#-Nonreltves under 18 in same hh	F1S93F	0.04	0.004	2.170	1.473	15227	0.003
Reading test formula score	F1TXRIR[d]	21.31	0.136	5.014	2.239	16304	0.061
Mathematics test formula score	F1TXMIR[d]	35.93	0.222	5.342	2.311	16270	0.096
Science test formula score	F1TXSIR[d]	13.80	0.092	5.341	2.311	16181	0.040
History/cit/geog test formla score	F1TXHIR[d]	19.11	0.099	4.816	2.194	16096	0.045

Mean	3.802	1.912
Minimum	1.708	1.307
Maximum	7.580	2.753
Standard deviation	1.574	0.390
Median	3.353	1.831

[a] This table is based on the original (1992-1993) 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 page 31 of section 3.1.2 for additional details about the sample numbers of the two releases.

[b] Standard error calculated taking into account the sample design.

[c] Standard error calculated under assumptions of simple random sampling.

[d] Although this table does not reflect the rescaling of first follow-up cognitive test items in the second follow-up, the correlation between the cognitive test items before and after the rescaling is 0.99.

**Table 3.3.1-5 NELS:88 first follow-up:
Mean design effects (DEFFs) and root design effects (DEFTs)
for student and dropout questionnaire data--full sample[a]**

Group	Mean DEFF	Mean DEFT
Students	3.858	1.923
Dropouts	4.713	1.999
Male[b]	3.370	1.797
Female	3.454	1.813
White	3.051	1.712
Black	3.615	1.827
Hispanic	3.555	1.755
Asian/Pacific Islander	2.765	1.627
American Indian/ Alaskan Native	2.415	1.442
Public schools	3.226	1.755
Catholic schools	2.668	1.535
Other private schools	6.650	2.421
Low SES	2.838	1.649
Middle SES	3.088	1.719
High SES	3.477	1.797
Urban	3.478	1.847
Suburban	3.475	1.799
Rural	2.668	1.578

[a] This table is based on the original (1992-1993) 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 page 31 of section 3.1.2 for additional details about the sample numbers of the two releases.

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

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

**Table 3.3.1-6 NELS:88 first follow-up:
Mean design effects (DEFFs) and root design effects (DEFTs)
for student and dropout questionnaire data--panel sample[a]**

Group	Mean DEFF	Mean DEFT
Students	3.802	1.912
Dropouts	4.705	1.997
Male[b]	3.456	1.817
Female	3.324	1.783
White	3.101	1.729
Black	3.804	1.867
Hispanic	2.643	1.591
Asian/Pacific Islander	2.758	1.609
American Indian/ Alaskan Native	2.066	1.362
Public schools	3.147	1.736
Catholic schools	2.619	1.513
Other private schools	6.529	2.391
Low SES	2.797	1.644
Middle SES	3.138	1.732
High SES	3.576	1.817
Urban	3.463	1.842
Suburban	3.412	1.788
Rural	2.634	1.571

[a] This table is based on the original (1992-1993) 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 page 31 of section 3.1.2 for additional details about the sample numbers of the two releases.

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

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

Table 3.3.1-7 NELS:88 first follow-up:
Standard errors and design effects, dropouts, full sample (*N*=1,043)[a]

Dropouts

Survey item (or composite variable)		Estimate	Design S.E.[b]	DEFF	DEFT	N	SRS S.E.[c]
R could not get along w/others	F1D6E	19.05	2.604	4.392	2.096	1000	1.243
R had no feeling of safety in school	F1D6K	11.41	2.142	4.535	2.129	1000	1.006
R had no feeling of belonging	F1D6P	24.97	3.230	5.563	2.359	1000	1.369
R dropped out because failing grades	F1D6R	42.10	3.506	5.038	2.245	1000	1.562
R had passing grade when last in school	F1D9	18.10	2.185	3.265	1.807	1015	1.209
Sts were in college prep/acad program	F1D16C	7.70	3.208	14.686	3.832	1015	0.837
Sts were in vocatnl/tech training	F1D16D	12.16	1.952	3.617	1.902	1015	1.026
Sts expect to finish college	F1D38	12.36	2.611	6.457	2.541	1027	1.027
At age 30 exp to be an employee	F1D39A	9.27	1.855	3.925	1.981	960	0.936
At age 30 exp to be a farmer	F1D39C	4.12	3.291	26.265	5.125	960	0.642
At age 30 exp to be a homemaker	F1D39D	3.01	0.828	2.255	1.502	960	0.551
At age 30 exp to be a manager	F1D39F	4.69	1.130	2.742	1.656	960	0.682
At age 30 exp to be in the military	F1D39G	3.61	0.652	1.172	1.083	960	0.602
At age 30 exp to be an operative	F1D39H	4.30	0.934	2.033	1.426	960	0.655
At age 30 exp to be a clergyman	F1D39J	7.45	2.708	10.201	3.194	960	0.848
At age 30 exp to be a school teacher	F1D39N	0.40	0.191	0.889	0.943	960	0.203
At age 30 exp to be a technician	F1D39P	2.90	0.600	1.227	1.108	960	0.542
At age 30 do not know what to be	F1D39S	15.16	1.735	2.244	1.498	960	1.158
Others in home speak spanish	F1D42	78.99	4.734	3.686	1.920	274	2.466
Live w/father in same house	F1D86A	31.16	2.558	3.084	1.756	1012	1.457
Live w/other adult male in hh	F1D86C	14.13	2.109	3.706	1.925	1012	1.095
Live with mother in same hh	F1D86D	69.97	2.814	3.810	1.952	1012	1.442
Live w/stepmother in same hh	F1D86E	2.66	0.635	1.576	1.255	1012	0.506
Live w/other adult female in hh	F1D86F	15.39	2.657	5.482	2.341	1012	1.135
Live with boy/girl friend	F1D86H	7.31	1.173	2.052	1.433	1012	0.809
Live with own children	F1D86I	18.42	2.448	4.031	2.008	1012	1.219
#-Sisters living in same hh	F1D87B	0.63	0.063	4.431	2.105	958	0.030
#-Grandparents in same hh	F1D87C	0.16	0.038	6.109	2.472	932	0.015
#-Relatives under 18 in same hh	F1D87D	0.19	0.030	1.056	1.028	934	0.029
#-Non relatives under 18 same hh	F1D87F	0.11	0.028	1.858	1.363	927	0.021

Mean	4.713	1.999
Minimum	0.889	0.943
Maximum	26.265	5.125
Standard deviation	4.953	0.860
Median	3.696	1.923

[a] This table is based on the original (1992-1993) 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 page 31 of section 3.1.2 for additional details about the sample numbers of the two releases.

[b] Standard error calculated taking into account the sample design.

[c] Standard error calculated under assumptions of simple random sampling.

Table 3.3.1-8 NELS:88 first follow-up:
Standard errors and design effects, dropouts, panel sample (*N*=765)[a]

Dropouts

Survey item (or composite variable)		Estimate	Design S.E.[b]	DEFF	DEFT	N	SRS SE[c]
R could not get alng w/others	F1D6E	20.05	3.228	4.784	2.187	737	[illegible]
R had no feeling of safety in school	F1D6K	12.12	2.648	4.845	2.201	737	[illegible]
R had no feeling of belonging	F1D6P	23.22	3.932	6.382	2.526	737	[illegible]
R dropped out because of failing grades	F1D6R	39.87	4.083	5.118	2.262	737	[illegible]
R had passng grades when last in school	F1D9	16.95	1.956	2.022	1.422	745	[illegible]
Sts were in college prep/acad program	F1D16C	8.43	4.084	16.035	4.004	743	[illegible]
Sts were in vocational/tech training	F1D16D	13.21	2.365	3.619	1.902	743	[illegible]
Sts expect to finish college	F1D38	11.84	3.177	7.300	2.702	756	[illegible]
At age 30 exp to be an employee	F1D39A	9.52	2.182	3.884	1.971	704	[illegible]
At age 30 exp to be a farmer	F1D39C	5.29	4.147	24.127	4.912	704	[illegible]
At age 30 exp to be a homemaker	F1D39D	2.20	0.786	2.016	1.420	704	[illegible]
At age 30 exp to be a manager	F1D39F	4.95	1.430	3.058	1.749	704	[illegible]
At age 30 exp to be in the military	F1D39G	3.54	0.788	1.277	1.130	704	[illegible]
At age 30 exp to be an operative	F1D39H	4.45	1.141	2.153	1.467	704	[illegible]
At age 30 exp to be a clergyman	F1D39J	6.73	2.772	8.611	2.934	704	[illegible]
At age 30 exp to be a school teacher	F1D39N	0.49	0.247	0.883	0.939	704	[illegible]
At age 30 exp to be a technician	F1D39P	2.92	0.678	1.142	1.068	704	[illegible]
At age 30 do not know what to be	F1D39S	15.03	2.012	2.228	1.493	704	[illegible]
Others in home speak spanish	F1D42	79.63	5.197	3.347	1.829	202	[illegible]
Live with father in same house	F1D86A	30.89	3.018	3.144	1.773	738	[illegible]
Live with other adult male in hh	F1D86C	14.28	2.502	3.769	1.941	738	[illegible]
Live with mother in same hh	F1D86D	68.29	3.366	3.856	1.964	738	[illegible]
Live with stepmother in same hh	F1D86E	2.83	0.780	1.631	1.277	738	[illegible]
Live with other adult female in hh	F1D86F	16.27	3.274	5.800	2.408	738	[illegible]
Live with boy/girl friend	F1D86H	7.62	1.394	2.033	1.426	738	[illegible]
Live with own children	F1D86I	18.90	2.932	4.133	2.033	738	[illegible]
#-sisters living in same household	F1D87B	0.62	0.077	5.433	2.331	696	[illegible]
#-grandparents in same household	F1D87C	0.17	0.047	6.252	2.500	674	[illegible]
#-relatives under 18 in same house	F1D87D	0.21	0.039	1.061	1.030	679	[illegible]
#-non relatves undr 18 in same hh	F1D87F	0.12	0.028	1.211	1.101	672	[illegible]

Mean	4.705	1.997
Minimum	0.883	0.939
Maximum	24.127	4.912
Standard deviation	4.748	0.862
Median	3.694	1.922

[a] This table is based on the original (1992-1993) 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 page 31 of section 3.1.2 for additional details about the sample numbers of the two releases.
[b] Standard error calculated taking into account the sample design.
[c] Standard error calculated under assumptions of simple random sampling.

**Table 3.3.1-9 NELS:88 second follow-up:
Standard errors and design effects, all respondents; full sample (*N*=19,220)**

All Students and Dropouts

Survey item (or composite variable)		Estimate	Design S.E.[a]	DEFF	DEFT	N	SRS S.E.[b]
There are many gangs in school	F2S7H	18.818	0.682	5.712	2.390	18761	0.285
I cut or skipped classes	F2S9B	2.956	0.073	4.610	2.147	18763	0.034
High school program - college prep	F2S12AB	35.860	0.679	3.796	1.948	18938	0.348
High school prgram - voc/tech prgms	F2S12AD	14.612	0.461	3.226	1.796	18938	0.257
Time watching TV during week	F2S35A[c]	78.539	0.520	2.633	1.623	16414	0.320
Being successful in line of work	F2S40A	98.733	0.156	3.699	1.923	19012	0.081
Level schl R's mother wants R cmplte	F2S42B	45.556	0.633	2.832	1.683	17532	0.376
Level school R anticipates completing	F2S43	30.215	0.610	3.245	1.801	18386	0.339
At age 30 R expects to be a manager	F2S64BF	5.777	0.251	2.105	1.451	18189	0.173
At age 30 R expects to be technician	F2S64BP	5.926	0.258	2.172	1.474	18189	0.175
I feel good about myself	F2S66A	93.523	0.291	2.401	1.549	17172	0.188
Luck more important than hard work	F2S66C	12.106	0.472	3.577	1.891	17082	0.250
Something always prevents success	F2S66F	25.916	0.578	2.968	1.723	17056	0.336
Plans hardly ever work out	F2S66G	21.750	0.564	3.177	1.782	16998	0.316
I do not have much to be proud of	F2S66L	15.860	0.471	2.823	1.680	16984	0.280
Chances R's life better than parents	F2S67K	60.872	0.651	3.005	1.734	16889	0.376
Number friends plan to attend college	F2S69E	48.259	0.750	3.931	1.983	17449	0.378
Relationship with fthr/mthr R's child	F2S79	25.365	2.195	3.510	1.873	1379	1.172
Amt earn/hour current/mst recent job	F2S91	5.472	0.027	2.848	1.688	11776	0.016
Amt earn from job R spends to go out	F2S92B	14.697	0.468	2.569	1.603	14706	0.292
Amt earn from job R spends on rent	F2S92D	3.876	0.269	2.844	1.687	14645	0.160
Last 2 yrs family memb in drug rehab	F2S96P	7.561	0.288	2.218	1.489	18690	0.193
Who decides if R can have job	F2S98C	57.361	0.701	3.143	1.773	15644	0.395
R's futr faml to be simlr to own faml	F2S100F	39.756	0.658	2.724	1.650	15069	0.399
English is native language	F2S107	10.732	0.747	11.118	3.334	19088	0.224
How well does R speak English	F2S109B	5.148	0.994	4.087	2.022	2020	0.492
Reading IRT-estimated number right	F22XRIRR	32.182	0.190	4.769	2.184	14176	0.087
Mathematics IRT-estmted nmbr right	F22XMIRR	46.859	0.290	5.559	2.358	14183	0.123
Science IRT-estimated number right	F22XSIRR	22.853	0.119	5.041	2.245	14080	0.053
Hist/Cit/Geo IRT-estmted nmbr right	F22XHIRR	34.279	0.102	4.917	2.217	14011	0.046

Mean	3.709	1.890
Minimum	2.105	1.451
Maximum	11.118	3.334
Standard deviation	1.685	0.369
Median	3.201	1.789

[a] Standard error calculated taking into account the sample design.
[b] Standard error calculated under assumptions of simple random sampling.
[c] Question asked on student questionnaire only.

Table 3.3.1-10 NELS:88 second follow-up:
Standard errors and design effects, all respondents; F2 panel sample (*N*=16,489)

All Students and Dropouts

Survey item (or composite variable)		Esti-mate	Design S.E.[a]	DEFF	DEFT	N	SRS S.E.[b]
There are many gangs in school	F2S7H	18.387	0.734	5.795	2.407	16142	0.305
I cut or skipped classes	F2S9B	2.897	0.081	5.063	2.250	16141	0.036
High school program - college prep	F2S12AB	37.986	0.754	3.933	1.983	16295	0.380
High school prgram - voc/tech prgms	F2S12AD	14.307	0.475	2.999	1.732	16295	0.274
Time watching TV during week	F2S35A[c]	78.433	0.532	2.410	1.552	14403	0.343
Being successful in line of work	F2S40A	98.791	0.170	3.955	1.989	16345	0.085
Level schl R's mother wants R cmplte	F2S42B	45.826	0.678	2.814	1.677	15197	0.404
Level school R anticipates completing	F2S43	30.671	0.625	2.919	1.709	15892	0.366
At age 30 R expects to be a manager	F2S64BF	5.515	0.255	1.960	1.400	15710	0.182
At age 30 R expects to be technician	F2S64BP	5.672	0.276	2.237	1.496	15710	0.185
I feel good about myself	F2S66A	93.518	0.293	2.122	1.457	14981	0.201
Luck more important than hard work	F2S66C	11.375	0.493	3.594	1.896	14908	0.260
Something always prevents success	F2S66F	25.341	0.608	2.908	1.705	14881	0.357
Plans hardly ever work out	F2S66G	21.263	0.612	3.320	1.822	14838	0.336
I do not have much to be proud of	F2S66L	14.963	0.484	2.729	1.652	14822	0.293
Chances R's life better than parents	F2S67K	61.002	0.702	3.055	1.748	14750	0.402
Number friends plan to attend college	F2S69E	50.206	0.809	3.954	1.989	15104	0.407
Relationship with fthr/mthr R's child	F2S79	26.631	2.642	3.880	1.970	1086	1.341
Amt earn/hour current/mst recent job	F2S91	5.459	0.030	3.114	1.765	10273	0.017
Amt earn from job R spends to go out	F2S92B	14.450	0.496	2.557	1.599	12848	0.310
Amt earn from job R spends on rent	F2S92D	3.386	0.238	2.215	1.488	12791	0.160
Last 2 yrs family memb in drug rehab	F2S96P	7.578	0.301	2.083	1.443	16102	0.209
Who decides if R can have job	F2S98C	56.753	0.721	2.897	1.702	13680	0.424
R's futr faml to be simlr to own faml	F2S100F	39.618	0.704	2.738	1.655	13217	0.425
English is native language	F2S107	8.814	0.649	8.600	2.933	16410	0.221
How well does R speak English	F2S109B	2.499	0.890	4.717	2.172	1451	0.410
Reading IRT-estimated number right	F22XRIRR	32.753	0.187	4.317	2.078	12718	0.090
Mathematics IRT-estmted nmbr right	F22XMIRR	47.593	0.291	5.169	2.273	12714	0.128
Science IRT-estimated number right	F22XSIRR	23.203	0.116	4.448	2.109	12631	0.055
Hist/Cit/Geo IRT-estmted nmbr right	F22XHIRR	34.583	0.101	4.428	2.104	12572	0.048

Mean	3.564	1.858
Minimum	1.960	1.400
Maximum	8.600	2.933
Standard deviation	1.366	0.332
Median	2.959	1.720

[a] Standard error calculated taking into account the sample design.
[b] Standard error calculated under assumptions of simple random sampling.
[c] Question asked on student questionnaire only.

Table 3.3.1-11 NELS:88 second follow-up:
Standard errors and design effects, all respondents; F1F2 panel sample (*N*=18,116)

All Students and Dropouts

Survey item (or composite variable)		Estimate	Design S.E.[a]	DEFF	DEFT	N	SRS S.E.[b]
There are many gangs in school	F2S7H	18.596	0.694	5.632	2.373	17700	0.292
I cut or skipped classes	F2S9B	2.931	0.076	4.997	2.235	17708	0.034
High school program - college prep	F2S12AB	36.665	0.706	3.835	1.958	17868	0.361
High school prgram - voc/tech prgms	F2S12AD	14.623	0.475	3.229	1.797	17868	0.264
Time watching TV during week	F2S35A[c]	78.707	0.528	2.592	1.610	15583	0.328
Being successful in line of work	F2S40A	98.694	0.165	3.788	1.946	17933	0.085
Level schl R's mother wants R cmplte	F2S42B	45.741	0.644	2.771	1.665	16585	0.387
Level school R anticipates completing	F2S43	30.104	0.618	3.153	1.776	17372	0.348
At age 30 R expects to be a manager	F2S64BF	5.767	0.261	2.156	1.468	17197	0.178
At age 30 R expects to be technician	F2S64BP	5.725	0.258	2.121	1.456	17197	0.177
I feel good about myself	F2S66A	93.560	0.279	2.105	1.451	16290	0.192
Luck more important than hard work	F2S66C	12.101	0.506	3.901	1.975	16206	0.256
Something always prevents success	F2S66F	25.957	0.579	2.823	1.680	16184	0.345
Plans hardly ever work out	F2S66G	21.779	0.572	3.098	1.760	16133	0.325
I do not have much to be proud of	F2S66L	15.577	0.467	2.673	1.635	16115	0.286
Chances R's life better than parents	F2S67K	61.023	0.667	2.997	1.731	16025	0.385
Number friends plan to attend college	F2S69E	48.775	0.772	3.934	1.983	16491	0.389
Relationship with fthr/mthr R's child	F2S79	25.138	2.313	3.551	1.884	1249	1.227
Amt earn/hour current/mst recent job	F2S91	5.463	0.028	2.063	1.750	11191	0.016
Amt earn from job R spends to go out	F2S92B	14.411	0.475	2.553	1.598	13958	0.297
Amt earn from job R spends on rent	F2S92D	3.465	0.219	1.993	1.412	13899	0.155
Last 2 yrs family memb in drug rehab	F2S96P	7.521	0.284	2.046	1.430	17642	0.199
Who decides if R can have job	F2S98C	57.199	0.702	2.990	1.729	14853	0.406
R's futr faml to be simlr to own faml	F2S100F	40.058	0.677	2.735	1.654	14331	0.409
English is native language	F2S107	10.071	0.768	11.732	3.425	18014	0.224
How well does R speak English	F2S109B	4.263	1.153	5.837	2.416	1792	0.477
Reading IRT-estimated number right	F22XRIRR	32.383	0.191	4.771	2.170	13668	0.088
Mathematics IRT-estmted nmbr right	F22XMIRR	47.059	0.289	5.345	2.312	13671	0.125
Science IRT-estimated number right	F22XSIRR	22.947	0.117	4.694	2.167	13574	0.054
Hist/Cit/Geo IRT-estmted nmbr right	F22XHIRR	34.381	0.103	4.803	2.191	13507	0.047

Mean	3.729	1.888
Minimum	1.993	1.412
Maximum	11.732	3.425
Standard deviation	1.844	0.405
Median	3.048	1.746

[a] Standard error calculated taking into account the sample design.
[b] Standard error calculated under assumptions of simple random sampling.
[c] Question asked on student questionnaire only.

**Table 3.3.1-12 NELS:88 second follow-up:
Mean design effects (DEFFs) and root design effects (DEFTs)
for student and dropout questionnaire data--full sample**

Group	Mean DEFF	Mean DEFT
All Respondents	3.709	1.890
Dropouts	2.929	1.690
Male[a]	3.080	1.724
Female	3.219	1.778
White	3.108	1.743
Black	2.959	1.690
Hispanic	2.830	1.647
Asian/Pacific Islander	2.690	1.621
American Indian/ Alaskan Native	3.276	1.686
Public schools	3.127	1.736
Catholic schools	2.594	1.577
Non-Catholic private schools	7.172	2.526
Low SES	2.936	1.681
Middle SES	2.529	1.574
High SES	3.963	1.950
Urban	3.868	1.925
Suburban	2.900	1.648
Rural	3.355	1.700

aSex categories are based on the composite sex variable.

Note: Each mean is based on 30 questionnaire items.

**Table 3.3.1-13 NELS:88 second follow-up:
Mean design effects (DEFFs) and root design effects (DEFTs)
for student and dropout questionnaire data--F2 panel sample**

Group	Mean DEFF	Mean DEFT
All Respondents	3.564	1.858
Dropouts	2.878	1.677
Male[a]	3.078	1.727
Female	3.208	1.759
White	3.101	1.733
Black	3.076	1.707
Hispanic	2.737	1.627
Asian/Pacific Islander	2.556	1.549
American Indian/ Alaskan Native	2.209	1.430
Public schools	2.934	1.681
Catholic schools	2.541	1.555
Non-Catholic private schools	7.301	2.577
Low SES	2.772	1.632
Middle SES	2.464	1.552
High SES	3.792	1.896
Urban	3.604	1.854
Suburban	2.936	1.686
Rural	3.074	1.639

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

Note: Each mean is based on 30 questionnaire items.

**Table 3.3.1-14 NELS:88 second follow-up:
Mean design effects (DEFFs) and root design effects (DEFTs)
for student and dropout questionnaire data--F1F2 panel sample**

Group	Mean DEFF	Mean DEFT
All Respondents	3.729	1.888
Dropouts	2.843	1.666
Male[a]	3.061	1.719
Female	3.209	1.768
White	3.015	1.713
Black	2.975	1.693
Hispanic	2.945	1.671
Asian/Pacific Islander	2.674	1.610
American Indian/ Alaskan Native	3.290	1.671
Public schools	3.148	1.735
Catholic schools	2.532	1.553
Non-Catholic private schools	7.368	2.591
Low SES	2.908	1.666
Middle SES	2.462	1.551
High SES	3.810	1.904
Urban	3.608	1.856
Suburban	3.005	1.707
Rural	3.556	1.714

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

Note: Each mean is based on 30 questionnaire items.

Table 3.3.1-15 NELS:88 second follow-up:
Standard errors and design effects, dropouts, full sample (*N*=2,028)

Dropouts

Survey item (or composite variable)		Esti-mate	Design S.E.[a]	DEFF	DEFT	N	SRS S.E.[b]
What year did R last attend school	F2D6Y	53.802	1.907	2.925	1.710	1999	1.115
What grade was R last in at school	F2D7	49.946	1.878	2.830	1.682	2006	1.116
Reason for leaving school	F2D9AD	15.312	1.289	2.445	1.564	1908	0.824
There are many gangs in school	F2D18H	28.201	1.861	3.281	1.811	1918	1.027
I cut or skipped classes	F2D19B	6.046	0.264	3.315	1.821	1912	0.145
High school program - college prep	F2D20C	5.030	0.558	1.248	1.117	1915	0.499
High school prgram - voc/tech prgms	F2D20D	14.878	1.540	3.586	1.894	1915	0.813
R enrlld in jr coll/voc programs	F2D23B	4.019	0.963	4.700	2.168	1955	0.444
Being successful in line of work	F2D36A	97.730	0.385	1.320	1.149	1976	0.335
Level schl R's mother wants R cmplte	F2D37B	30.854	1.910	3.184	1.784	1862	1.070
Level school R anticipates completing	F2D38	11.042	1.299	3.223	1.795	1876	0.724
At age 30 R expects to be a manager	F2D40AD	8.637	0.892	1.969	1.403	1953	0.636
At age 30 R expects to be technician	F2D40AO	9.050	0.940	2.097	1.448	1953	0.649
Amt earn/hour current/mst recent job	F2D45K	5.611	0.076	2.221	1.490	1534	0.051
Amt earn from job R spends to go out	F2D47B	9.453	1.024	1.860	1.364	1518	0.751
I feel good about myself	F2D57A	91.491	1.008	2.341	1.530	1794	0.659
Luck more important than hard work	F2D57C	18.906	1.879	4.117	2.029	1788	0.926
Something always prevents success	F2D57F	42.633	1.948	2.773	1.665	1787	1.170
Plans hardly ever work out	F2D57G	34.341	1.742	2.400	1.549	1783	1.125
I do not have much to be proud of	F2D57L	21.810	1.575	2.598	1.612	1786	0.977
Chances R's life better than parents	F2D58K	52.523	2.077	3.095	1.759	1789	1.181
Number friends plan to attend college	F2D59E	13.463	1.371	3.143	1.773	1948	0.773
Relationship with fthr/mthr R's child	F2D69	32.167	3.343	3.693	1.922	721	1.740
Events occrd in R's family last 2 yrs	F2D80L	13.352	1.164	2.285	1.512	1951	0.770
Last 2 yrs family memb in drug rehab	F2D80P	10.583	0.980	1.982	1.408	1953	0.696
Who decides if R can have job	F2D81C	84.902	2.011	3.821	1.955	1211	1.029
R's futr faml to be simlr to own faml	F2D82F	47.811	2.513	3.045	1.745	1203	1.440
English is native language	F2D89	13.010	1.695	5.100	2.258	2009	0.751
How well does R speak English	F2D91B	6.604	2.995	4.348	2.085	299	1.436

Mean	2.929	1.690
Minimum	1.248	1.117
Maximum	5.100	2.258
Standard deviation	0.921	0.272
Median	2.801	1.674

[a] Standard error calculated taking into account the sample design.
[b] Standard error calculated under assumptions of simple random sampling.

Table 3.3.1-16 NELS:88 second follow-up:
Standard errors and design effects, dropouts, F2 panel sample (*N*=1,512)

Dropouts

Survey item (or composite variable)		Esti- mate	Design S.E.[a]	DEFF	DEFT	N	SRS S.E.[b]
What year did R last attend school	F2D6Y	56.860	2.215	2.978	1.726	1489	1.284
What grade was R last in at school	F2D7	49.785	2.202	2.902	1.703	1496	1.293
Reason for leaving school	F2D9AD	14.155	1.468	2.525	1.589	1424	0.924
There are many gangs in school	F2D18H	28.239	2.210	3.451	1.858	1432	1.190
I cut or skipped classes	F2D19B	5.839	0.313	3.471	1.863	1428	0.168
High school program - college prep	F2D20C	5.261	0.626	1.127	1.061	1433	0.590
High school prgram - voc/tech prgms	F2D20D	16.437	1.872	3.656	1.912	1433	0.979
R enrlld in jr coll/voc programs	F2D23B	3.459	0.963	4.066	2.016	1464	0.478
Being successful in line of work	F2D36A	97.694	0.475	1.479	1.216	1477	0.391
Level schl R's mother wants R cmplte	F2D37B	30.818	2.258	3.343	1.828	1398	1.235
Level school R anticipates completing	F2D38	9.709	1.084	1.883	1.372	1405	0.790
At age 30 R expects to be a manager	F2D40AD	9.177	1.068	1.995	1.413	1458	0.756
At age 30 R expects to be technician	F2D40AO	8.433	1.003	1.899	1.378	1458	0.728
Amt earn/hour current/mst recent job	F2D45K	5.630	0.097	2.529	1.590	1157	0.061
Amt earn from job R spends to go out	F2D47B	8.970	1.227	2.109	1.452	1144	0.845
I feel good about myself	F2D57A	91.183	1.203	2.407	1.551	1337	0.775
Luck more important than hard work	F2D57C	17.018	1.998	3.774	1.943	1335	1.029
Something always prevents success	F2D57F	43.891	2.226	2.680	1.637	1332	1.360
Plans hardly ever work out	F2D57G	35.823	2.202	2.805	1.675	1330	1.315
I do not have much to be proud of	F2D57L	21.097	1.682	2.262	1.504	1331	1.118
Chances R's life better than parents	F2D58K	52.094	2.463	3.248	1.802	1336	1.367
Number friends plan to attend college	F2D59E	13.064	1.459	2.735	1.654	1459	0.882
Relationship with fthr/mthr R's child	F2D69	34.498	4.132	4.080	2.020	540	2.046
Events occrd in R's family last 2 yrs	F2D80L	13.007	1.430	2.640	1.625	1461	0.880
Last 2 yrs family memb in drug rehab	F2D80P	10.850	1.242	2.332	1.527	1462	0.813
Who decides if R can have job	F2D81C	85.079	2.137	3.169	1.780	881	1.200
R's futr faml to be simlr to own faml	F2D82F	47.699	3.000	3.149	1.775	873	1.690
English is native language	F2D89	13.023	1.650	3.605	1.899	1500	0.869
How well does R speak English	F2D91B	6.376	3.758	5.157	2.271	218	1.655

Mean	2.878	1.677
Minimum	1.127	1.061
Maximum	5.157	2.271
Standard deviation	0.847	0.254
Median	2.707	1.645

[a] Standard error calculated taking into account the sample design.
[b] Standard error calculated under assumptions of simple random sampling.

Table 3.3.1-17 NELS:88 second follow-up:
Standard errors and design effects, dropouts, F1F2 panel sample (*N*=1,837)

Dropouts

Survey item (or composite variable)		Estimate	Design S.E.[a]	DEFF	DEFT	N	SRS S.E.[b]
What year did R last attend school	F2D6Y	55.902	1.945	2.778	1.667	1810	1.167
What grade was R last in at school	F2D7	51.284	1.928	2.702	1.644	1816	1.173
Reason for leaving school	F2D9AD	15.184	1.356	2.473	1.573	1732	0.862
There are many gangs in school	F2D18H	27.603	1.942	3.278	1.811	1737	1.073
I cut or skipped classes	F2D19B	5.953	0.267	3.045	1.745	1733	0.153
High school program - college prep	F2D20C	5.369	0.606	1.256	1.120	1737	0.541
High school prgram - voc/tech prgms	F2D20C	15.307	1.594	3.404	1.845	1737	0.864
R enrlld in jr coll/voc programs	F2D23B	3.303	0.798	3.531	1.879	1771	0.425
Being successful in line of work	F2D36A	97.596	0.416	1.321	1.149	1791	0.362
Level schl R's mother wants R cmplte	F2D37B	31.098	2.007	3.177	1.782	1690	1.126
Level school R anticipates completing	F2D38	10.080	1.016	1.936	1.391	1700	0.730
At age 30 R expects to be a manager	F2D40AD	8.859	0.965	2.039	1.428	1768	0.676
At age 30 R expects to be technician	F2D40AO	8.522	0.927	1.949	1.396	1768	0.664
Amt earn/hour current/mst recent job	F2D45K	5.618	0.080	2.278	1.509	1391	0.053
Amt earn from job R spends to go out	F2D47B	9.628	1.136	2.041	1.429	1376	0.795
I feel good about myself	F2D57A	91.267	1.071	2.339	1.529	1625	0.700
Luck more important than hard work	F2D57C	19.036	2.102	4.647	2.156	1621	0.975
Something always prevents success	F2D57F	44.550	2.040	2.729	1.652	1620	1.235
Plans hardly ever work out	F2D57G	35.558	1.879	2.491	1.578	1617	1.190
I do not have much to be proud of	F2D57L	21.624	1.657	2.621	1.619	1618	1.023
Chances R's life better than parents	F2D58K	52.575	2.192	3.124	1.767	1621	1.240
Number friends plan to attend college	F2D59E	13.105	1.283	2.559	1.600	1770	0.802
Relationship with fthr/mthr R's child	F2D69	31.577	3.566	3.796	1.948	645	1.830
Events occrd in R's family last 2 yrs	F2D80L	13.030	1.269	2.515	1.586	1770	0.800
Last 2 yrs family memb in drug rehab	F2D80P	10.661	1.074	2.145	1.465	1771	0.733
Who decides if R can have job	F2D81C	84.634	2.179	3.998	1.999	1095	1.090
R's futr faml to be simlr to own faml	F2D82F	48.615	2.681	3.136	1.771	1090	1.514
English is native language	F2D89	13.086	1.684	4.545	2.132	1823	0.790
How well does R speak English	F2D91B	6.439	3.204	4.584	2.141	269	1.497

Mean	2.843	1.666
Minimum	1.256	1.120
Maximum	4.647	2.156
Standard deviation	0.872	0.259
Median	2.590	1.609

[a] Standard error calculated taking into account the sample design.
[b] Standard error calculated under assumptions of simple random sampling.

Catholic schools. In NELS:88, the average design effect is 2.70; in High School and Beyond, it was 3.60 for the sophomores and 3.58 for the seniors.

The smaller design effects in the NELS:88 base year may reflect the somewhat smaller cluster size used in the later survey. The High School and Beyond 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.

3.3.2 First Follow-Up Standard Errors and Design Effects

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.

Selection of First Follow-Up Items. 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 High School and Beyond analyses of standard errors/design effects. This criterion resulted in the selection of ten questions, seven which were used in both the NELS:88 base year and High School and Beyond standard error/design effects analysis and three 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 in order to ensure that variables that were important to analysts, thus likely to receive considerable use, were represented. Using this criterion, four cognitive test scores, specifically the IRT-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 multi-level math and reading test forms.

The remaining 16 variables were selected randomly from the pool of remaining critical items. The selection process occurred using 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 from by the size of their means and a systematic sample of 16 items was selected from the sorted list of items.

For dropouts, the starting point for selecting the variables for standard error/design effect calculations was to use items that overlapped 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.

Results. Standard errors and design effects were calculated for each of the 30 items for the sample as a whole and for selected subgroups. The subgroups were based on the respondent's school status (student/dropout), sex, race and ethnicity, school type (public, Catholic, and other private), socioeconomic status (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.

The individual item standard errors, design effects (DEFF) and root design effects (DEFT) for all respondents are presented along with summary statistics in Tables 3.3.1-3 (full sample) and 3.3.1-4 (panel sample). Tables 3.3.1-5 and 3.3.1-6 present corresponding summary design effects for the subgroups.

Individual item standard errors, design effects and design effect summary statistics for dropouts are presented in Tables 3.3.1-7 (full sample) and 3.3.1-8 (panel sample). No subgroup analyses were conducted for the dropouts because the resulting sample sizes would have been quite small. Individual item standard errors and design effects by subgroups are presented in the *NELS:88 First Follow-Up Final Technical Report*.[16]

As expected, the design effects in the first follow-up are somewhat higher than those of the base year. This is a result of

[16] Ingels S.J., Scott L.A., Rock D., Pollack J., Rasinski K.; Washington D.C.: NCES, 1994.

the subsampling procedures used for the first follow-up; students who were found to be attending schools with a small number of base year sample students were undersampled in the first follow-up. Tables 3.3.1-5 and 3.3.1-6 show that subgroups also have larger design compared to those in the base year. Table 3.3.1-2 presents base year design effects for 12 subgroups defined similarly to those in Tables 3.3.1-5 and 3.3.1-6. For 11 of the twelve 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 one exception is students from private schools. While having the highest average design effect (as they did in the base year analysis), these students show a lower average design effect in the first follow-up survey (full sample, 6.65; panel sample, 6.53) than in the base year survey (8.80).

Both average design effects for the first follow-up survey were larger than the average design effect of 2.88 obtained for the base year HS&B Sophomore Cohort. The direction of this difference held for 10 of the 11 subgroups comparable across the first follow-up and HS&B. Catholic school students are the exception. The average first follow-up design effect for Catholic school students is lower than the average HS&B Catholic school student design effect (first follow-up: full sample, 2.67, panel sample, 2.62; HS&B, 3.60). While the first follow-up design effect for private school students was higher than in HS&B, the difference is small (first follow-up: full sample, 6.65, panel sample, 6.53; HS&B, 6.22); in fact it is the smallest of the differences in average design effects between the two surveys.

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, that is, it increases design effects. This is so because subsampling 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 HS&B base year. Comparatively more subsampling was done in the NELS:88 first follow-up, which has 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 are 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) 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.

3.3.3 Second Follow-Up Standard Errors and Design Effects

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.

Selection of Second Follow-Up Items. 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 High School and Beyond 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. This criterion was applied in order to ensure that variables that are important to analysts, thus likely to have a higher frequency of use, were represented. Using this criterion, four cognitive test scores were selected--the IRT-estimated number right scores for mathematics, English, science, and social studies. Although several test score composites were available, the IRT-estimated number right scores were used because they compensate for guessing and omitted items. The IRT scores have also been equated across the multi-level math and reading test forms.

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), socioeconomic status (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 just the panel respondents weighted by F2PNLWT, and the third using just the respondents in the first and second follow-up panel sample weighted by F2F1PNWT.

The individual item standard errors, design effects (DEFF) and root design effects (DEFT) for all respondents are presented along with summary statistics in Tables 3.3.1-9 (full sample) and 3.3.1-10 (panel sample), and 3.3.1-11 (first/second follow-up panel sample). Tables 3.3.1-12, 3.3.1-13, and 3.3.1-14 present corresponding summary design effects for the subgroups.

Individual item standard errors, design effects and design effect summary statistics for dropouts are presented in Tables 3.3.1-15 (full sample) and 3.3.1-16 (panel sample), and 3.3.1-17 (first/second follow-up panel sample). As in the first follow-up analysis, no subgroup analyses were conducted for the dropouts because the resulting sample sizes would have been quite small. Individual item standard errors and design effects by subgroups are presented in the forthcoming *NELS:88 Second Follow-Up Sample Design Report*.

The design effects in the second follow-up are lower than those in the first follow-up (for both the full sample and the panel) but higher than those in the base year. Tables 3.3.1-12, 3.3.1-13,and 3.3.1-14 show that, for the most part, the second follow-up design effects for subgroups are also larger than those obtained for similar subgroups in the base year (see Table 3.3.1-2 for comparison). For 11 of the twelve subgroups in the full sample, and for 10 of the twelve subgroups in the panel samples, the second follow-up survey average design effects are larger than those for the base year survey. The exceptions are students from Catholic and other private schools, although the design effect for other private schools remains the highest of all the second follow-up subgroups for the full and panel samples.

As mentioned earlier, the tendency in longitudinal studies is for design effects to lessen over time because of dispersion of the sample members from the original clusters. However, subsampling introduces additional variability into the weights with an attendant loss in sample efficiency. The second follow-up design effects are probably larger than the base year design effects because of the subsampling in the first follow-up. They are most likely smaller than the design effects of the first follow-up because of sample dispersion between the first and second follow-ups. When the NELS:88 second follow-up design effects are compared to those from the HS&B first follow-up of the sophomore cohort a remarkable similarity is found. DEFF is 3.709 for the full sample NELS:88 second follow-up data, and 3.589 for the equivalent HS&B first follow-up data. DEFT is 1.890 for NELS:88 and 1.837 for

HS&B.

3.3.4 Design Effects and Approximate Standard Errors

Researchers who do not have access to software for computing accurate estimates of standard errors can use the mean design effects presented in Tables 3.3.1-2 (for base year data) 3.3.1-5 and 3.3.1-6 (for first follow-up data), and 3.3.1-12, 3.3.1-13 and 3.3.1-14 (for second follow-up data) to approximate the standard errors of statistics based on the NELS:88 data. 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):

$$SE = DEFT \times (p(1-p)/n)^{1/2} \qquad (1)$$

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:

$$SE = DEFT \times (Var/n)^{1/2} \qquad (2)$$

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 presented in the preceding section make it clear that the design effects and root design effects vary considerably by subgroup. It is therefore important to use the mean DEFT for the relevant subgroup in calculating approximate standard errors for subgroup statistics.

Standard error estimates may be needed for subgroups that are not tabulated here. 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 generally be less affected by clustering than larger subgroups.) Estimates for Hispanic males, for example, will generally have smaller design effects than the corresponding estimates for all Hispanics or all males. For this reason, it will usually be conservative to use the subgroup mean DEFT to approximate standard errors for estimates concerning a portion of the subgroup. This rule applies only when the variable used to subdivide a subgroup crosscuts schools. Sex 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 the 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:

$$\mathrm{Var}(b-a) < \mathrm{Var}(b) + \mathrm{Var}(a) \qquad (3)$$

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. It follows from equation (3) that Var(a) + Var(b) can be used in place of Var(b-a) with conservative results.

A final rule of thumb is that more complex estimators show smaller design effects than simple estimators.[17] Thus, correlation and regression coefficients tend to have smaller design effects than subgroup comparisons, and subgroup comparisons have smaller design effects than means. This implies that it will be conservative to use the mean root design effects presented here in calculating approximate standard errors for complex statistics, such as multiple regression coefficients. The procedure for calculating such approximate standard errors is the same as with simpler estimates: first, a standard error is calculated using the formula for data from a simple random sample; then, the 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 (e.g., $NEWWGT= ((1/DEFF)*(F2QWT_i/(\Sigma F2QWT_i/N)))$ for second follow-up full sample data), and use this new weight to deflate the obtained sample size to take into account the inefficiencies due to a sample design that is a departure from a simple random sample. Using this procedure, statistics calculated by a statistical program such as 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 analysis that assumes the data were collected from

[17] Kish, L., and Frankel, M. (1974). Inference from complex samples. *Journal of the Royal Statistical Society: Series B (Methodological), 36, 2-37.*

a simple random sample. The analyst applying this correction procedure should carefully examine the statistical software he or she is using, and assess whether the program treats weights in such a way as to produce the effect described above.

3.4 Additional Sources of Nonobservational Error

Analysis of survey error is important for understanding the potential bias in making inferences from an obtained sample to a population. Sampling errors occur because the data are collected from a sample rather than a census of the population. Sampling error analyses for NELS:88 (documenting standard errors of measurement and design effects for key variables) were presented earlier in this chapter (see section 3.3). In this section, other sources of nonobservational error are discussed.

Nonobservational error results from measurements not being taken from a portion of the population.[18] Several factors comprise nonobservational error, including nonresponse biases caused by unit and item nonresponse and undercoverage. Nonresponse is readily quantified. While many data quality factors are difficult to measure in the non-experimental context of large-scale survey administration, NELS:88 offers the possibility of comparing reports from multiple sources, thereby permitting some approximate but useful validity parameters. Below, we discuss two kinds of nonobservational error in the NELS:88 second follow-up: undercoverage and nonresponse.

3.4.1 Biases Caused by Undercoverage of Special Populations

Undercoverage of Non-English Speakers. There is significant undercoverage in the NELS:88 data of the portion of the language minority population that is more severely limited in English proficiency (LEP) or non-proficient (NEP) in English. This undercoverage is most severe for the base year questionnaire data, and for test results from all waves of NELS:88. Undercoverage bias will affect estimates for LEPs and NEPs, but will also affect certain estimates for racial-ethnic subgroups that have large numbers of LEPs and NEPs when individuals in these groups generally differ in a relevant characteristic from other non-LEP/NEP Asians, Hispanics or others. Although, for example, Hispanics and Asians were selected at a higher than normal rate in the base year, have been disproportionately retained in subsequent follow-ups, and have been added to the cohort as their eligibility status was found to have changed, significant numbers of Asian, Hispanic and other LEPs

---

[18] Groves, R. M., *Survey Errors and Survey Costs*. New York: John Wiley and Sons, 1989, page 11.

were excluded from the base year sample.[19]

Specifically, among the total number of eighth-grade students enrolled in the 1,052 fully participating base year schools, 1.9 percent of the potential sample (3,831 of 202,966) were excluded by their schools for reasons of a language barrier to participation. Had no students been excluded for language reasons, the NELS:88 baseline would have included an additional 532 students. All of these students would be classifiable as LEPs or NEPs; 270 of these excluded students were Hispanics, 175 were Asians, and the remaining 87 language-excluded eighth-grade students were of another race/ethnicity (neither Hispanic nor Asian). Some 24,599 students (out of 26,432 sample members) participated in the base year, and of these participants, 642 were classified either by self-report or teacher report as of limited English proficiency. If one counts as LEP all students reported as LEP by either source, then just over half of the LEPs in the potential sample were captured by the base year sample design and contributed data to the base year. (If one uses the more stringent criterion of counting only those so identified by both sources--self-report and teacher--or counts only those identified by teachers, then less than half of the potential LEPs are represented in the base year data.)

Initially in the first follow-up and then in the second follow-up, two measures were adopted to increase coverage of students with limited English language proficiency. First, eligibility rules were modified so that the number of LEPs obtained through sample freshening would be maximized. The modified eligibility rules were applied also to the sample of base year ineligibles in the first follow-up and to the ineligibles in the second follow-up followback study of excluded students. Second, base year and first follow-up ineligibles who had gained sufficient proficiency to complete survey forms in the first and second follow-ups were added to the cohort. Students with a language barrier who were reclassified were administered the student questionnaire in Spanish or English, or the dropout questionnaire (in English or Spanish) if they were school-leavers. Enrollment status data was gathered for those students who were classified as being still unable to complete the NELS:88 survey forms.

LEPs who Entered the Sample through Freshening. Substantial

[19] Of course, elements excluded from the sampling frame are not accounted for by sample weighting so that population estimates from the data file fall appropriately short of full 1987-88 eighth-grade enrollment figures. Nevertheless, such exclusions limit one's ability to describe in an unbiased way special populations of interest, such as all dropouts, all language minority students, and so on. Some examples of this potential for bias may serve to underline the need for caution in the use of the language minority student data.

numbers (236 total in the first and second follow-up rounds of freshening) of limited English proficient students entered NELS:88 through the freshening process. LEPs are, of course, disproportionately present in the population of students who fall behind the modal progression through school. While, by the most generous count (that is, self-report or teacher report), only 2.6 percent (or, weighted, 2.3%) of the base year respondents were LEPs, around 17 percent of the freshening sample in first follow-up were classified by their schools as LEPs (176 out of 1,060). Virtually all of the LEP students selected in the freshening process were retained for the first follow-up.[20] Similarly, 69 of the 288 (24 percent) students identified in the second follow-up freshening process were classified by their schools as LEP; 60 (87 percent) of these LEP students were added to the NELS:88 cohort during the second follow-up.[21]

As noted above, eligibility rules were modified in the first follow-up to reduce the likelihood that LEP students would be excluded in the sample freshening process. With support from the Office of Bilingual Education and Minority Language Affairs (OBEMLA), the student questionnaire was translated into Spanish for both the 1990 and 1992 rounds; because a translation of the cognitive tests was not feasible, students completing the Spanish questionnaire were not pressed to attempt to complete the test component.

LEPs who Entered the Sample through Studies of Excluded Students. The same modified eligibility rules were applied retroactively to a sample of base year language-excluded students in the first and second follow-ups. Language-excluded students whose English proficiency status had changed such that they were able to complete the survey forms were administered the English-language version of the student or dropout questionnaire. Although cognitive test data were not collected for this group in the first follow-up, as many of these students as possible (45, or 34 percent) were tested in the second follow-up in 1992. The 532 students who would have been chosen for the base year except for

[20] Three had to be excluded because they had physical or mental disabilities that precluded their participation, and eleven were temporarily ineligible (out of scope for the first follow-up because though in the country at the time of freshening, they were outside the country at the time of data collection). The other 158 entered the first follow-up sample.

[21] Of the remaining 9 LEPs identified for freshening in the second follow-up, 5 were out of the country at the time of data collection, 3 had mental or physical disabilities that precluded their participation, and one spoke a language other than Spanish and could not complete survey instruments in English.

language barriers to their participation were represented (with appropriate adjustment to their weights) in the first follow-up base year ineligibles study by 204 individuals; of these, 131 were found to be eligible (of which 118 participated) and were included in the NELS:88 cohort in the second follow-up. The eligibility of the remaining 73 language-excluded students was reassessed in the second follow-up followback study of excluded students (FSES); of these 73, 22 were found to be eligible and 19 (86.4 percent) participated.[22]

LEP students added to the cohort through the freshening process appear on this data file. First follow-up data for base year language ineligibles who have become eligible did not appear on the initial 1991 public release file, but have been integrated into the first follow-up files and will appear in subsequent combined releases of NELS:88 data (1994 electronic codebook release). Since it was not necessary to exclude any freshened students for language reasons in the first follow-up and only one student was excluded in the second follow-up, and because cases representing about 74 percent of the base year language exclusions became eligible in either the first or second follow-up, the net effect of these additions to the data is to substantially reduce undercoverage of current and former limited English-proficient students. However, bias is at best but modestly reduced for the cognitive test data because some of the freshened LEP students and second follow-up FSES eligibles did not complete the cognitive tests, and none of the first follow-up reclassified base year excluded students completed the test battery. Data users should take these potential biases into account in their analyses.

Undercoverage of Students with Disabilities. There is significant undercoverage in the NELS:88 data of that portion of the special education population that is most severely mentally or physically disabled. Undercoverage bias may also affect certain estimates for racial or gender subgroups that have large numbers of students in the excluded category. (Our data show, for example, that blacks and males are disproportionately represented in the class of students excluded owing to mental disability). Coverage of this population was improved in the first follow-up by the fact that in the base year ineligibles study, nine of the 23 students excluded because of physical barriers to participation, and 140 of the 322 students who had been excluded because of mental barriers to participation, were reclassified as eligible. Similarly, 49 of the previously ineligible sample members were found to be eligible in the second follow-up followback study of excluded students; of these 49 excluded students, 44 had been previously excluded due to mental disability and 5 for physical limitations. However, it is our sense that very few of these students actually "changed"

[22] Of these 73 excluded students, 40 were screened and determined to be ineligible, 21 had moved out of the country, and 12 remained unscreened.

substantially between rounds; rather, most reclassifications reflected the process of taking a second look at students at the margin between eligible and ineligible, and aggressively pursuing status information from their special education teachers, information that would permit a more accurate assessment to be made of their ability to complete at least the student questionnaire. Overwhelmingly, the reclassified students would appear to be those with learning disabilities or emotional disturbances, rather than the mentally retarded. Hence students with severe or profound impairments are not represented in the NELS:88 data.

Estimates based on the members of the ineligibles sample are also subject to limitations. By and large, the NELS:88 samples of eligible and ineligible language-excluded students, when combined, provide excellent population coverage. However, for the severely physically and mentally disabled populations, there are two potential sources of exclusion in addition to school-level classification as ineligible. These further sources of undercoverage are 1) exclusion of schools (special purpose schools for students with disabilities were excluded from the base year sampling frame), and 2) the exclusion of ungraded classrooms in what was by definition a sample of eighth graders.

Test Score Undercoverage of Dropouts. Data users are reminded that no special nonresponse adjusted weight was created for cases with a completed questionnaire but without a cognitive test. As in the base year, cognitive test completion rates were sufficiently high that such a weight was not needed. Rates of test completion among in-school sample members were 96.5 percent in the base year and 94.1 percent in the first follow-up, with a decrease to 76.6 percent in the second follow-up.

However, the high overall rate of test completion for students does not apply to dropouts. While 91 percent of identified dropouts provided questionnaire data in the first follow-up, cognitive tests were completed by only half of the sample members who completed a full or abbreviated dropout questionnaire.[23] In the second follow-up, 88 percent of the dropouts provided questionnaire data but only 42 percent completed a cognitive test. This low rate of test completion is attributable to the high percentage of questionnaires that were administered by telephone, as well as to the strategy of obtaining questionnaire data only rather than accepting a refusal from a dropout or alternative completer unwilling to take the cognitive test. Of course, base year test

---

[23] According to the first follow-up design, dropouts administered the abbreviated or modified dropout questionnaires (28% of the dropout sample) were not asked to complete the cognitive test battery; for these sample members only the standard classification variables and a number of key items that differentiate the in-school and out-of-school populations are available for analysis.

score data are available for most of the individuals for whom first and/or second follow-up test results were not obtained. **It would be inadvisable to, for example, draw conclusions about test score gains between 1988 and 1990 or between 1990 and 1992 for dropouts as a separate group, given the amount of 1990 and 1992 test data that are missing.**

3.4.2 Unit and Item Nonresponse

Unit Nonresponse. Unit nonresponse occurs when an individual respondent (such as a student, school administrator, or teacher) declines to participate, or when the cooperation of a school cannot be secured. In the base year, an analysis of school-level nonresponse suggested 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 the quality of the student sample is small (for details, see the *Base Year Sample Design Report*). School nonresponse has not been 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 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.

The effect of student-level nonresponse within the responding schools was not assessed in the base year, although males, blacks, and Hispanics tended to be nonparticipants more often than females, whites or Asians. Note that NELS:88 weights adjust for unit nonresponse.

Item Nonresponse. As noted above, sampling and coverage errors are two key components of total survey error. Sampling error is quantified through the standard errors and design effects for key variables. There are other sources and types of nonobservational error, including estimate error or bias associated with unit (individual) nonresponse and item nonresponse. In addition to its role as a potential source of bias, item nonresponse also has the effect of diminishing the number of observations that can be used in calculating statistics from affected data elements and thus increases sampling variances. Since item nonresponse is an important potential and uncorrected source of data bias, it is necessary to measure its impact so that analysts can properly take potential response biases into account when developing their analysis plans. NCES's standard asserts that total weighted nonresponse for an item (unit nonresponse multiplied by item nonresponse) should not exceed 30 percent; items that exceed that standard have been noted in the codebook. This section reports specifically on nonsampling measurement error as a function of item nonresponse.

Item nonresponse occurs when a respondent fails to complete certain items on the survey instrument. While bias associated with unit nonresponse has been controlled by making adjustments to case weights, item nonresponse has generally not been compensated for in the NELS:88 student component data set. There are three exceptions to this generalization.

The first exception is machine editing, through which certain nonresponse problems are rectified for some items by imposing inter-item consistency, particularly by forcing logical agreement between filter and dependent questions. For example, the missing response to a filter question can often be inferred if dependent questions have been answered. Because the edited files were used in the nonresponse analysis reported below, this adjustment to item nonresponse is reflected in the results of the analysis.

The second exception is that some key classification variables have been constructed in part from additional sources of information when questionnaire data are missing. Data from school records (for example, student sex or race/ethnicity as given on the sampling roster) or other respondent sources (for example, the parent questionnaire) have been used to replace missing data. See section 7.2.3 for further information on constructed classification variables. Because composite variables were not included in the nonresponse analysis, this adjustment of missing data is *not* reflected in the statistics reported below.

The third exception is the language series filter question. Base year and first follow-up data were imported into the second follow-up files in order to resolve missing cases; in particular, to identify respondents who should have legitimately skipped the dependent items in the language series. This adjustment to nonresponse is reflected in the item statistics reported below.

A further point to note is that there may be some hidden nonresponse in the NELS:88 base year and first follow-up questionnaire data that is impossible to quantify. This is the case because many questions use a "mark all that apply" format. While such a format results in slightly less burden to the respondent, it also makes it impossible to distinguish between a negative response and nonresponse. This conflation of negative response and nonresponse creates the potential for nonresponse biases that cannot be measured and thus cannot become the basis for precise warnings to users about the limitations of data. In the second follow-up most "mark all that apply" formats were changed to an explicit "yes" or "no" response for each subitem. This change in format did not entirely eliminate the nonresponse problem; the data show that for long lists of subitems, respondents seem to mark only one type of response ("yes" for those subitems that apply). To minimize item nonresponse for these questions, response patterns were analyzed and inferences made about missing responses.

A final point is that unit nonresponse is a further source of missing item data--nonparticipating students complete no questionnaire items. Weights accommodate student nonresponse by projecting questionnaire data to the full population, with appropriate adjustments for defined subgroups. However, nonresponse-adjusted weights cannot compensate for the bias that arises if nonrespondents and respondents would have answered the questionnaire differently. Hence "total response" should be thought of as the survey (unit) response rate times the item response rate. (For example, given a cross-sectional weighted student response rate of 91 percent, and an item response rate of 88 percent, total response would be 80 percent.)

Two main objectives guide the following item nonresponse analysis. One objective is to quantify mean student questionnaire nonresponse overall as well as nonresponse for the entire in-school sample on key variables that appeared on the student questionnaire. A second objective is to describe nonresponse patterns in terms of item characteristics. In order to realize the first objective, average nonresponse rates were calculated for each item. To fulfill the second objective, nonresponse was measured as a function of three item characteristics: 1) position in the questionnaire; 2) topic; and 3) whether the item was contingent on a filter.

Population and Data File Definitions.

Definition 1: "Item"

For purposes of this analysis, "item" refers to each data element or variable. For a question composed of multiple subparts, each subpart eliciting a distinct response is counted as an item for item nonresponse purposes. (Thus, a single question that poses three subquestions is treated as three variables.)

Definition 2: "Response Rate"

NCES standards stipulate that item response rates (Ri) "are to be calculated as the number of respondents for whom an in-scope response was obtained (i.e., the response conformed to acceptable categories or ranges), divided by the number of completed interviews for which the question (or questions if a composite variable) was intended to be asked.":

$$Ri = \frac{\text{weighted \# of respondents with in-scope responses}}{\text{weighted \# of completed interviews for which question was intended to be asked}}$$

In-scope responses were considered to be valid answers (including a "don't know" response when this was a legitimate response option). Out-of-scope responses were multiple responses to items requiring only a single response, refusals, and missing responses.

Definition 3: "Analysis Populations"

Item nonresponse analysis population--student questionnaire. All students who completed any form of the questionnaire, regardless of whether they completed the test.

Definition 4: "Student and Dropout Questionnaire Data File"

The public use data file with machine-edited, weighted data was used as the basis for the analysis. Nonresponse rates of composite and other constructed variables and test data were not examined in this analysis.

Definition 5: "Nonresponse"

For the student and dropout questionnaires several numerical reserved codes were used to categorize nonresponse. The reserved codes and definitions appear below. The first three--reserved codes 6, 7 and 8--define out-of-scope or illegitimate nonresponse, and were used as the basis for this nonresponse analysis.

6 = Multiple Response. For an item that required one response only, the respondent marked more than one response, and the multiple response could not be resolved.

7 = Refused Critical Item. Respondent was unwilling to answer the question at the time of the questionnaire administration and upon nonresponse follow-up by survey administrators.

8 = Missing. The response datum is illegitimately missing. That is, a datum that should be present for this respondent is missing. Data elements not appearing on the abbreviated or modified student or dropout questionnaires were considered as illegitimately missing.

9 = Legitimate Skip. The response datum is legitimately missing. That is, owing either to responses to preceding filter questions or to other respondent characteristics, data for this item should not be present for this respondent. Responses under reserved code 9 were not included in the nonresponse analysis.

DK = Don't Know. "Don't Know" is often used as a nonresponse code. In the NELS:88 data set, "Don't Know" is embedded as a legitimate response category in some of the questionnaire items. For purposes of this analysis, "Don't Know" was not classified as a nonresponse.

Item-Level Nonresponse. Table 3.4.2-1 shows descriptive statistics for item nonresponse for the student questionnaire overall and for items grouped into categories depending upon their position in the questionnaire, the topic they addressed, and whether they were part of a skip or filter pattern.

The mean item nonresponse rate for the NELS:88 second follow-up student questionnaire is 12.1 percent, compared to 4.7 percent on the base year instrument and 7.0 percent in the first follow-up.

A special factor influencing item nonresponse rates in the first and second follow-up was the administration of different versions of the student questionnaire. The two versions of the questionnaires differed in the number of questions being asked of respondents. For purposes of item response analyses, questions not appearing on the abbreviated or modified student questionnaire were treated as if they were intended to be asked of all participating sample members. This was done so that the total impact on estimation of missing information--whether the information was missing by design, or by respondent omission or error--could be assessed. Hence, completed abbreviated or modified interviews were included in the denominator of the item response formula used in this analysis. Out of the 17,192 student questionnaire respondents, only 1,489 or 8.7 percent completed either a modified or abbreviated questionnaire; most of these completed an instrument modified for telephone administration. Appendix L contains a complete list of the items excluded from the versions of the student questionnaire used for telephone administration and refusal conversion.

Item-Level Nonresponse by Item Placement and Characteristic

Item Nonresponse by Position in Questionnaire. The pattern of item nonresponse by position in the questionnaire is similar to that experienced in the NELS:88 base year and first follow-up. Average item nonresponse in the first third of the instrument is 8.2 percent (base year, 3.5 percent; first follow-up, 4.3 percent). For the middle questions, average item nonresponse rises to 10.5 percent (base year, 3 percent; first follow-up, 8.5 percent), with a sharper rise in mean item nonresponse in the last third of the questionnaire (17.7 percent, as compared to 7.5 percent in the base year and 8.2 percent in the first follow-up). Because there are many high nonresponse outliers in the middle third of the first follow-up student questionnaire, comparisons of the middle and last third of that questionnaire mask the effect on the data of the progressive increase in nonresponse as one approaches the end of the survey administration session and poorer readers and less motivated respondents face difficulties in completing the instrument. In the second follow-up, time available for questionnaire completion for some respondents was further compressed due to the obligations of older students (for example, work study arrangements, mid-morning transfer to another campus for vocational education classes, and class tests that seniors did not

Table 3.4.2-1
Percent nonresponse on the student questionnaire by various item characteristics

Domain	Average	Standard Deviation	Minimum	Maximum	Number of Items
Overall Position	12.07	11.28	.00	71.32	564
First Third	8.20	7.29	.60	29.31	189
Second Third	10.52	7.53	.00	38.85	193
Last Third	17.73	15.17	.00	71.32	182
Topic (in order of appearance in the questionnaire)					
School Experiences	8.19	6.58	.60	29.31	236
Future Plans	11.18	8.85	.00	38.85	133
Opinions, Attitudes	5.24	10.35	1.66	42.40	83
Money and Work	10.96	3.86	1.38	19.21	17
Family	11.12	7.15	.00	29.06	52
Language Use	31.56	21.05	.74	71.32	43
Filtered					
No	7.24	5.04	.00	20.71	277
Yes	16.73	13.48	.00	71.32	287

wish to miss). Although the second follow-up student questionnaire was no doubt somewhat too long for some respondents to complete (the number of items rose from a total of 475 in the first follow-up to 564 in the second follow-up), nonresponse in the final third of the instrument is comparable to that in HS&B. Even in the last section of the questionnaire applicable to all respondents (the final section covered language use, which most respondents could legitimately skip out of after response to one item asking for native language), total response--item response of about 88 percent and unit response of about 91 percent--yields an 80 percent total response rate, well within the range specified in NCES statistical standards.

Item Nonresponse by Topic. The NELS:88 questionnaires have been organized topically in each wave; each section represented a different theme. Table 3.4.2-2 lists the topical sections in the second follow-up instrument in the order in which they appeared in the questionnaire. Nonresponse rates for the second follow-up, compared with those from the base year and first follow-up, are depicted side by side, with topics listed in the order of their appearance in the second follow-up questionnaire. For purposes of

Table 3.4.2-2
Percent item nonresponse by topical area[a]

Topic	F2 Non-Response	F1 Non-Response (Position)		BY Non-Response (Position)	
(2) School Experiences	8.19	4.5	(1)	6.9	(7)
(3) Future Plans	11.18	3.9	(2)	2.5	(5)
(4) Opinions, Attitudes	5.24	6.7	(4)	1.6	(4)
(5) Money and Work	10.96	10.9	(6)	0.9	(6)
(6) Family	11.12	8.9	(7)	3.4	(3)
(7) Language Use[b]	31.56	34.2	(3)	5.0	(2)

[a] This table is based on the original (1992-1993) 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. Additional details about the sample numbers of the two releases are on page 31 of section 3.1.2, under the subheading "Subsampling the Eighth-Grade Cohort and Freshened Sophomore Samples."
[b] Questionnaire sections on Address Information (section 1) and Background have not been included in this analysis.

comparison, the relative locations of the thematic section in the base year and first follow-up instruments are also indicated.

Given its position in a questionnaire that is nearly twice as long as the base year student questionnaire and more than a hundred questions longer than the first follow-up instrument, it is not surprising that items in the language use section have far higher nonresponse rates than in the first follow-up or the base year. Since most respondents skipped out of this question series, data were collected from only a small subset of the student population. Nevertheless, the respondent population for this series is particularly of interest for policy reasons and the apparent increase from the modest 5 percent nonresponse in the base year is dramatic.

Three related factors contribute to high item nonresponse in the language section. First, illegitimate skips at the filter carry missing data forward into dependent items. (The relevant file-building convention--operative in NLS-72, HS&B and the NELS:88 base year as well--is that items missing on a filter are also coded as missing on the dependent series.) Second, progressive subsetting of the relevant population (the filter is followed by two further filters) increases the proportion of missings even while their absolute number remains relatively stable. At the same time, the ambiguous nature of the missings renders the extent of true nonresponse for any given data element impossible to ascertain. The third factor is the generally poor language skills of the targeted population. The operation of these factors may be illustrated by reference to the data.

The first question in the language section--F2S107, which asks what the respondent's native language (language first spoken) was--is a crucial filter. Because of its critical nature and the nonresponse problem experienced in the first follow-up, this item was designated as critical in the second follow-up; however, this did not ameliorate the problem as had been hoped. Those answering "English" were skipped to the request for written permission to collect a high school transcript--that is, skipped out of the language section entirely. Those answering with a language other than English are given no instructions, though it is implicit that they should go to question 108, rather than skipping to the transcript request. In the original data (prior to cross-wave editing in which base year and first follow-up responses were drawn upon to "clean" many of the second follow-up missings on F2S107), students failed to respond at the filter question. These missings, carried into the dependent series, increased nonresponse substantially. As further filters reduce the relevant population to smaller subsets, the missings are carried to subsequent filter and dependent questions, where they loom as an ever larger proportion of the total. For example, by the time we reach the subsequent filter at F2S110A, the unambiguously specified population for defining the subset is 2,194 cases, while the number of ambiguous missings is only 434. This creates a very high and partly spurious nonresponse rate in the dependent items to F2S110A (F2S110B and F2S110C). Similar problems were experienced in other sections of the questionnaire, notably in series that asked about military service and about respondent's child or children.

Item Nonresponse by Dependence on a Filter Question. As is clear from the discussion of problems in the language section above, skip patterns contributed significantly to second follow-up item nonresponse. As noted in Table 3.4.2-1, questions that were not dependent on previous filter questions had a nonresponse rate of 7.2 percent, while those that were dependent had a rate of 16.7 percent. In the base year, the nonresponse rate for filtered questions was 5.8 percent, and 4.5 percent for unfiltered; in the first follow-up, the nonresponse rate was 12.7 percent for filtered questions and 5.6 percent for unfiltered after invoking base year data for cross-wave editing (nonresponse for filtered items was 14.45 percent prior to such cleaning). Even though eighth graders as a group are generally thought to be less able to deal with skips than high school students, they apparently had far less difficulty with routing instructions than students (largely, the same students) in the first and second follow-ups. HS&B base year and sophomore cohort first follow-up skip pattern item nonresponse reflects much lower rates than NELS:88 first and second follow-ups, perhaps because they used far fewer filter questions. The pattern for the NELS:88 second follow-up is similar to the NLS-72 base year, which likewise used many filter items.

Several factors contributed to the substantial increase in the level of item nonresponse in the NELS:88 first and second follow-ups over levels of filtered item nonresponse registered in the base year. First, on the basis of field test results, the most difficult filter series was made a critical item (subject to retrieval) in the base year and thus had the benefit of interviewer critical item edits. Second, formats were less crowded and routing arrows were employed to help students follow skips, when the "skip to" item appeared on the same page as the filter (the predominate case--by design--in the base year). Third, no abbreviated or modified questionnaires were employed in the base year data collection.

In contrast, the NELS:88 first and second follow-ups did not use the HS&B approach of minimizing the number of filter questions and making virtually all filter items critical, and therefore subject to field edit and retrieval. Nor was the base year strategy of using a combination of critical item status and, where the routing could be contained within a single visual format such as a page or facing pages, the use of routing arrows employed. There were eight major skips in the first follow-up questionnaire, and 25 in the second follow-up student questionnaire. Of these second follow-up skips, only seven were designated as critical items. In addition, the first follow-up questionnaires did not consistently give "go to" instructions for students who were not to follow the skip. This omission abetted respondent error in items such as F1S13, F1S54, F1S58, F1S84, and F1S95. These differences in questionnaire design account for much of the dramatically higher rate of missings associated with filter-dependent items in NELS:88 first follow-up as contrasted to HS&B and NELS:88 base year; "go to" instructions were consistently included in the second follow-up instruments. However, just over one percent of first follow-up respondents and 10.8 percent of second follow-up respondents were administered abbreviated or modified instruments, resulting in some items being skipped by design. While first follow-up nonresponse resulting from the use of abbreviated versions of these questionnaires had a minor effect on response rates overall, the impact was proportionally more for filtered subsets of the population. The impact of abbreviated questionnaires in the second follow-up was of somewhat greater magnitude and was more evenly distributed among subpopulations.

<u>Student Survey Item-Level Nonresponse by Critical Items</u>. Since a complete edit with data retrieval for *all* missing items would be prohibitively expensive for most surveys, the conventional strategy is to identify a subset of "key" or "critical" items for each survey instrument which, if not answered, triggers an attempt to recontact the respondents to obtain the missing data.

The average second follow-up nonresponse rate for the 69 critical student items is 3.3 percent (unweighted, 2.9 percent), compared with an average of 2.7 percent on 42 critical items (if one outlier that performed uniquely--BYS31B--is excluded) and 2.6

percent on 50 critical items in the first follow-up. As a further point of comparison, the HS&B sophomore cohort first follow-up questionnaire in 1982 had approximately 40 critical data points with 3.7 as the mean percentage of missing data.

Weighted nonresponse on key items ranged from zero percent to nearly 13 percent. The item nonresponse rates for each of the critical items in the student questionnaire are shown in Table 3.4.2-3. Note that the table provides both weighted and unweighted item nonresponse rates for the critical items, as both are useful. From a methodological perspective, the quality of given items can best be assessed with raw data, since nonresponse adjustments generalize data to nonrespondents as well as respondents. And, since Asians and Hispanics were oversampled, and typically carry smaller weights, while transfer students carry very high weights, interactions with subgroup responding characteristics can introduce distortions. On the other hand, from an analytic point of view, the weighted data provide a more meaningful item response rate, since the analyst is interested in population estimates and the extent of item nonresponse with application of the final weights has been taken into account.

Overall, the second follow-up had a high rate of unit (student) response. Cross-sectionally, around 93 percent of students and 88 percent of dropouts participated overall, while 96 percent of the in-school portion of the longitudinal cohort of eighth graders participated. These rates match the achieved 93 percent base year completion rate and the 94 percent student completion rate (91 percent for dropouts) in the first follow-up. Weighted response rates were 91 percent for students cross-sectionally in 1990 and 93 percent for the panel (1988 participants who also participated in 1990 as students).[24] The weighted completion rate for dropouts was 91 percent. While markedly higher than the base year and first follow-up, a reasonable rate of item nonresponse (the overall nonresponse rate based on weighted data is 12.1 percent) was achieved. For a number of format and other questionnaire design

[24] While weighted response rates are slightly higher than raw response rates in the base year and for first follow-up dropouts, the weighted response rate is lower than the raw completion rate for the first and second follow-up student questionnaires. This largely reflects the effects of subsampling in the first follow-up, with lower completion rates for groups with higher weights (for example, a 20% subsample was taken of the transfer students, and transfers participated at a substantially lower rate than other students).

Table 3.4.2-3
Nonresponse for critical items in the student questionnaire

Item Number	Weighted Percent Not Responding	Unweighted Percent Not Responding
F2S6A	0.60	0.41
F2S11A	1.19	0.94
F2S12A	1.17	0.98
F2S25A1	3.59	3.25
F2S25A2	3.92	3.47
F2S25B1	3.20	3.06
F2S25B2	3.74	3.37
F2S25C1	4.57	4.24
F2S25C2	5.26	4.50
F2S25D1	4.14	3.83
F2S25D2	4.72	4.09
F2S25E1	3.71	3.33
F2S25E2	4.14	3.62
F2S25F1	4.18	3.78
F2S25F2	4.08	3.42
F2S40A	1.10	0.91
F2S40B	1.14	0.92
F2S40C	1.14	0.95
F2S40D	1.14	0.95
F2S40E	1.19	0.99
F2S40F	1.23	1.05
F2S40G	1.24	1.06
F2S40H	1.26	1.05
F2S40I	1.30	1.10
F2S40J	1.37	1.17
F2S40K	1.34	1.09
F2S40L	1.43	1.13
F2S40M	1.35	1.11
F2S40N	1.26	1.02
F2S40O	1.23	0.99
F2S44A	2.95	2.69
F2S44B	2.88	2.65
F2S44C	3.26	3.05
F2S44D	4.00	4.00
F2S44E	3.36	3.08
F2S44F	4.62	4.32
F2S60A	1.42	1.27
F2S64A	6.88	6.31
F2S64B	6.00	5.56
F2S76	1.66	1.40
F2S86A	1.38	1.08
F2S86BMO	7.38	6.74
F2S86BYR	8.46	7.47

Note: For a list of the actual questions, refer to Appendix L.

Table 3.4.2-3 (cont.)
Nonresponse for critical items in the student questionnaire

Item Number	Weighted Percent Not Responding	Unweighted Percent Not Responding
F2S95	5.63	4.87
F2S96A	2.75	2.34
F2S96B	2.82	2.42
F2S96C	3.02	2.60
F2S96D	2.96	2.58
F2S96E	3.14	2.73
F2S96F	3.03	2.65
F2S96G	3.01	2.59
F2S96H	3.02	2.62
F2S96I	2.96	2.55
F2S96J	2.95	2.55
F2S96K	2.99	2.58
F2S96L	3.01	2.62
F2S96M	3.50	3.12
F2S96N	3.03	2.67
F2S96O	3.18	2.74
F2S96P	3.01	2.65
F2S96Q	2.98	2.60
F2S107	0.74	0.66
F2S108A	8.29	6.67
F2S108B	8.80	7.12
F2S108C	8.95	7.21
F2S108D	8.58	6.80
F2S108E	13.00	8.90

Note: For a list of the actual questions, refer to Appendix L.

reasons, filter questions appeared to work less efficiently in the first and second follow-ups than in the base year, and contributed to the higher item nonresponse--to both genuine nonresponse and to an undeterminable amount of artifactual nonresponse. The average nonresponse rate for critical items in the student questionnaire is around 3.3 percent. In terms of questionnaire length, while nonresponse is noticeably high in the last section of the questionnaire, it is attributable to both a long instrument and to the "nested" skips within the section, which causes very high item nonresponse within the subitems of the nested pattern and drives the average item nonresponse in the section above the NCES standard. Total nonresponse based on weighted data is around 20 percent (with unit nonresponse at 9 percent and mean item nonresponse for responding units at 12 percent).

3.4.3 Observational Error: The Quality of Responses

Observational errors, deviations of the answers of respondents from their true values, stem from a complex set of factors, including the respondent's knowledge and motivation in interaction with the instrument, the adequacy of the instrument, and its mode of administration.[25] As Fetters, Stowe and Owings (1984, p. vii) note, "the quality of student questionnaire data depends on both the nature of the questions asked and the characteristics of the student who provides the answer."[26] This observation, though drawn from the analysis of questionnaire results, is equally applicable to cognitive test data.

Cognitive Test Battery Reliabilities. Results of psychometric analyses of the second follow-up cognitive test battery--including score means and standard deviations, reliabilities (coefficient alpha), and standard errors of measurement--will be presented in the NELS:88 Second Follow-Up Psychometric Report. For details on base year test differential item functioning, item statistics and other characteristics of the base year test data, see the *Psychometric Report for the NELS:88 Base Year Test Battery*.[27] Also, the results of psychometric analyses of the first follow-up test battery are reported in the *NELS:88 First Follow-Up Final Technical Report.*[28]

Base Year Quality of Student Responses. Kaufman, Rasinski, Lee, and West assessed the reliability and validity of NELS:88 base year student data.[29] Their report examined the correspondence between parent and student responses to similar items, the consistency among student responses to related items, and the internal consistency reliability of scalable survey responses. Their general conclusions were that NELS:88 data exhibited a high degree of consistency and accuracy. Users of the base year data files may wish to consult the full report for further information

---

[25] R. Groves, 1989, *Survey Errors and Survey Costs*, page 11.

[26] Fetters, W.B., Stowe, P.S., and Owings, J.A. 1984. *High School and Beyond: Quality of Responses of High School Students to Questionnaire Items*. Washington, D.C.: U.S. Department of Education, NCES.

[27] Rock, D.A., and Pollack, J.M.; Washington D.C.: NCES, 1991.

[28] Ingels S. J., Scott L.A., Rock D., Pollack J., Rasinski K.; Washington D.C.: NCES, 1994.

[29] Kaufman, P., Rasinski, K., Lee, R. and West, J. 1991. *Quality of the Responses of Eighth-Grade Students in NELS:88*. Washington, DC, U.S. Department of Education, NCES 91-487.

on the quality of particular data elements, scales and constructs. When using models that incorporate a provision for measurement error, analysts may wish to consider using the reported validity coefficients as adjustment factors. Spencer, Frankel, Ingels, Rasinski, and Tourangeau analyzed high nonresponse items in the base year student questionnaire in order to determine the relationship between item nonresponse and student characteristics.[30] They found that item nonresponse was higher among males than females, and among blacks and Hispanics than among whites and Asians. Summary data on quality of base year student responses are provided in Appendix P.

Quality of Responses to the First and Second Follow-Up Student Questionnaires. At this time, extensive data quality analyses have not been conducted for the first or second follow-ups. However, quality of response analyses were conducted for the HS&B tenth- and twelfth-grade data of 1980 by Fetters, Stowe and Owings. Given that HS&B in 1980 was a similar survey conducted under comparable conditions and with comparable populations, some of the broader conclusions drawn from the HS&B analyses are likely to apply to the data in NELS:88.

The HS&B analyses examined student questionnaire data validity, using the parent questionnaire data and high school transcripts as the standard. Reliability coefficients were estimated from twin data.

Fetters, Stowe and Owings found a number of student characteristics to be associated with differences in data reliability and validity. High school seniors provided better quality data than did sophomores, and female students provided slightly better information than did males. White students provided better quality data than did Hispanic or black students, and students with high cognitive test scores provided better data than did students with low scores on the HS&B tests. In general, Fetters, Stowe and Owings found that contemporaneous and factually-oriented items were more reliable and valid than subjective and retrospective items.

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