
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 States 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 NELLS: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, Shio-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

Foreword	i
A Note on Data Use and Confidentiality	iii
Acknowledgements	iv
I. Introduction	1
1.1 Organization of the Data User's Manuals	1
1.2 Overview	2
1.2.1 NCES's National Education Longitudinal Studies Program	2
1.2.2 The National Longitudinal Study of the 1970s: NLS-72	2
1.2.3 High School and Beyond of the 1980s: HS&B	3
1.3 The National Education Longitudinal Study of 1988: Overview	6
1.3.1 NELS:88 Study Objectives	7
1.3.2 Base Year Study and Sample Design	9
1.3.3 First Follow-Up Core Study and Sample Design	12
1.3.4 Second Follow-Up Core Study and Sample Design	13
1.3.5 Second Follow-Up Design Enhancements	15
1.4 NELS:88 Sponsors	16
1.4.1 Sample Supplements and Augmentations	16
1.4.2 Instrument Supplements	16
1.5 NELS:88 Data and Documentation	17
1.5.1 Base Year Data Tapes and Documentation	18
1.5.2 First Follow-Up Data Files and Documentation	18

1.5.3	Second Follow-Up Electronic Codebook on CD-ROM and Documentation	20
II. Data Collection Instruments		21
2.1	Instrument Development	21
2.2	Survey Instruments and Content Coverage	22
2.2.1	Student Questionnaire and Cognitive Tests	22
2.2.2	Dropout Questionnaire	24
2.2.3	New Student Supplement	26
2.2.4	Early Graduate Supplement	26
2.2.5	Adapting Questionnaires for Telephone Administration	26
III. Sample Design and Implementation; Survey Error Assessment		28
3.1	NELS:88 Sample Design	28
3.1.1	Base Year Sample Design	28
3.1.2	First Follow-Up Sample Design	29
3.1.3	Second Follow-Up Sample Design	34
3.2	Calculation of Weights	42
3.2.1	Calculation of Base Year Sample Weights	42
3.2.2	Calculation of First Follow-Up Sample Weights	44
3.2.3	Calculation of Second Follow-Up Weights	48
3.3	Standard Errors and Design Effects	55
3.3.1	Base Year Standard Errors and Design Effects	56
3.3.2	First Follow-Up Standard Errors and Design Effects	85
3.3.3	Second Follow-Up Standard Errors and Design Effects	88
3.3.4	Design Effects and Approximate Standard Errors	90
3.4	Additional Sources of Nonobservational Error	92

3.4.1	Biases Caused by Undercoverage of Special Populations	92
3.4.2	Unit and Item Nonresponse	97
3.4.3	Observational Error: The Quality of Responses .	109
IV.	Data Collection	111
4.1	Base Year Data Collection	111
4.1.1	Base Year Pre-Data Collection Activities	111
4.1.2	Base Year Cohort Data Collection Activities . .	112
4.1.3	Base Year Data Collection Results	113
4.2	First Follow-Up Data Collection	113
4.2.1	First Follow-Up Pre-Data Collection Activities .	114
4.2.2	First Follow-Up Cohort Data Collection Activities	116
4.2.3	First Follow-Up Dropout Survey	117
4.2.4	First Follow-Up Survey of Base Year Ineligible Students	119
4.2.5	First Follow-Up 1990 and 1988-90 Panel Data Collection Results	121
4.3	Second Follow-Up Data Collection	121
4.3.1	Second Follow-Up Pre-Data Collection Activities	123
4.3.2	Second Follow-Up Cohort Data Collection Activities	126
4.3.3	Second Follow-Up Student Survey and Cognitive Tests	126
4.3.4	Second Follow-Up Dropout Survey	130
4.3.5	School Effectiveness Study	131
4.3.6	Followback Study of Excluded Students (FSES) . .	133
4.3.7	Second Follow-Up Data Collection Results	134

V.	Data Control and Preparation	143
5.1	On-Site Editing and Retrieval	143
5.2	Monitoring and Receipt Control	143
5.3	In-House Editing and Coding	143
5.4	Data Capture and Archival Storage	144
VI.	Data Processing of the Student Questionnaires	145
6.1	Machine Editing	145
6.2	Data File Preparation	147
6.3	CD-ROM Electronic Codebook	148
VII.	Guide to the Data Files, Documentation and CD-ROM Electronic Codebooks	149
7.1	Basics for Analyses: Second Follow-Up Questionnaire and Sample Indicators, Statistical Weights and Use of Statistical Packages	153
7.1.1	Questionnaire/Sample Flags Included on Magnetic Tape and ECB Releases	153
7.1.2	Packaged Statistical Programs	165
7.2	Content and Organization of the Data Files	165
7.2.1	Identification Codes	167
7.2.2	The Student Survey Instruments	167
7.2.3	Composite Variables	170
7.3	Guide to the NELS:88 Codebooks	173
7.3.1	Hardcopy Codebooks in NELS:88 Data User's Manuals	174
7.3.2	The NELS:88 Electronic Codebook System (ECB)	178

Appendices

Appendix A: NELS:88 Sources of Contextual Data: Parent, Teacher, School Administrator, Transcript, and Course Offerings Components

- Appendix B:** NELS:88-Related Data Files Available from the National Center for Education Statistics
- Appendix C:** National Center for Education Statistics, Longitudinal and Household Studies Branch, NELS:88 Publications
- Appendix D:** Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors: Analytical Implications of Design Differences Between the Studies
- Appendix E:** NELS:88 Second Follow-Up Item Overlap with NELS:88 Base Year and First Follow-Up; and with HS&B and NLS-72
- Appendix F:** NELS:88 Second Follow-Up Followback Study of Excluded Students: Screeners
- Appendix G:** Public Use Record Layout for NELS:88 Second Follow-Up Data Tape
- Appendix H:** NELS:88 Student Data Weights, Flags, and Composite Variables (BY, F1, and F2)
- Appendix I:** Guidelines for Using SAS with NELS:88 Second Follow-Up Student Data
- Appendix J:** NELS:88 Second Follow-Up Student Questionnaire Codebook
- Appendix K:** NELS:88 Second Follow-Up Student Questionnaire and New Student Supplement
- Appendix L:** Critical Items, Abbreviated Questionnaire Items, and Refusal Conversion Items from the Second Follow-Up Student and Dropout Questionnaires and New Student Supplement
- Appendix M:** NELS:88 Second Follow-Up Content Areas: Student, Dropout and School Components
- Appendix N:** NELS:88 Second Follow-Up Content Areas: Student, Teacher and Parent Components
- Appendix O:** Glossary of NELS:88 Terms
- Appendix P:** Selected Measures of NELS:88 Base Year Data Quality
- Appendix Q:** NELS:88 Third Follow-Up (1994) Questionnaire

V. Data Control and Preparation

This chapter describes the procedures used to control student data before transforming responses from second follow-up questionnaires into a data file. Several procedures were implemented to prepare these documents for data capture, including monitoring the receipt of completed questionnaires, editing completed questionnaires, retrieving missing data, and preparing the documents for archival storage. Data preparation activities spanned the entire length of the NELS:88 second follow-up student survey, beginning with tracing and securing school cooperation, through monitoring and machine editing, and ending with the preparation of public use data files.

5.1 On-Site Editing and Retrieval

For student and dropout questionnaires (including the new student supplement), the first data control and preparation activity was editing questionnaires and retrieving missing information. Interviewers conducted on-site editing of the student and dropout questionnaires, giving special attention to the respondents' answers for all critical items. A list of critical items can be found in Appendix L.

If the response to one or more of the critical items was missing, undecipherable, or had multiple categories marked when only one response was permitted, the interviewer privately pointed out the problem to the respondent. If the sample member indicated that he or she had chosen not to answer the question, the interviewer marked a "no retrieval" response for the item. The "no retrieval" responses were later used during the machine editing process to assign a "refused" response to the critical items.

5.2 Monitoring and Receipt Control

Once the questionnaires, cognitive tests, and new student supplements were collected, each student/dropout questionnaire was reviewed for completeness and to confirm that the ID numbers were correct. A final disposition code was assigned to each student and dropout indicating whether test data, questionnaire data, or a combination of the two were completed by the sample member. These outcomes were recorded in a microcomputer-based Survey Management System (SMS).

5.3 In-House Editing and Coding

The next step was to edit the confidential locator pages for legibility and remove the pages from the questionnaire. In the student questionnaire respondents were asked to provide the names and locations of the two postsecondary institutions they were most likely to attend after high school. This information was coded using the standard Interagency Postsecondary Education Data System (IPEDS) codes. (IPEDS codes are available only on the restricted

use files.)

5.4 Data Capture and Archival Storage

Data entry for the student questionnaire and cognitive tests was performed through an optical mark reading procedure by Questar Data Systems, Inc. The new student supplements and dropout questionnaires were not optically scanned but were converted to machine readable form using conventional key-to-disk methods. All cognitive tests were photographed onto microfilm for archival storage.

VI. Data Processing of the Student Questionnaires

In each round of the study, data processing activities began with sample selection and continued through receipt control, machine edit, and the preparation of public and restricted use data files and user documentation. Data processing activities varied little among the base year, first follow-up and second follow-up. This chapter describes the post-processing that was carried out to prepare the data for final release. This chapter concludes with an introduction to the electronic codebooks (ECBs) that have been created for NELS:88 data.

6.1 Machine Editing

Conventions for editing, coding, error resolution, and documentation adhered as closely as possible to the procedures and standards previously established for HS&B and NLS-72.

Detection of out-of-range codes was completed during scanning or data entry for all questions except those permitting an open-ended response. The scanning contractor converted the student data to machine-readable form and supplied a raw data tape to NORC. Because of their small number, the new student supplements were not scanned, but were hand-keyed. After receipt of all scanned and keyed data, sequenced machine editing and visual inspection of the output began. The tasks performed included: resolving inconsistencies between filter and dependent questions, supplying the appropriate missing data codes for questions left blank, detecting illegal codes and converting them to missing data codes and investigating inconsistencies or contradictions in the data. Frequencies and crosstabulations for each variable were inspected before and after these steps to verify the accuracy and appropriateness of the automated machine editing processes.

Inconsistencies between filter and dependent questions were resolved in the machine editing process. In most instances, dependent questions that conflicted with the skip instructions of a filter question contained data that, although possibly valid, were superfluous. For instance, respondents sometimes indicated "no" to a filter question and then continued to answer "no" to subsequent dependent items. When a filter question indicated that a subsequent question(s) should have been skipped, the dependent questions were set to the value "legitimate skip", with one exception. In the exception, if the dependent questions were answered in a manner that was inconsistent with the filter but consistent across the dependent items, the filter was back edited (changed) to agree with the dependent responses. If a multiple response or no answer was given to a filter question, the question was assigned the appropriate reserved code (see below) and all subsequent questions that might have been skipped were processed as if the respondent should have answered them.

The frequency with which responses were recoded to legitimate

skip for each skip pattern was closely monitored. Frequency distributions of responses before and after editing were inspected. All filter questions and their respective dependent items were displayed in crosstabulations so that staff could verify the accuracy of the recoding.

After improperly answered questions were converted to blanks, the student data were passed through a second step in the editing program that supplied the appropriate reserved codes for blank questions. Where a value was not provided by the respondent, a reserved code fills the field. These reserved codes and their meanings are as follows:

6=MULTIPLE RESPONSE
7=REFUSED⁷
8=MISSING
9=LEGITIMATE SKIP

When the legitimate response of a variable filled more than one column of space, the right-hand column contained one of the above codes and the remainder of the columns were filled with "9"s.

Critical items (those deemed most critical to data analyses) followed a somewhat different machine editing process. Data collection procedures instructed field interviewers to mark the retrieval oval beside each critical item in the questionnaire if an attempt was made to retrieve missing or invalid data from a respondent. The edit program then used these fields to set corresponding blank data to "refused." Since their purpose was to determine the correct reserved codes, retrieval variables are not present on the final data file. If a critical item was left blank, was not a legitimate skip, and an attempt was made to retrieve the missing data, the item was coded as "8" (missing). If a filter was coded "7" (refused), all subsequent questions that might have been skipped were processed as if the respondent should have answered each item. Filters that were coded "6" (multiple response) or "8" (missing) were handled in the same manner.

Items with unusually high nonresponse or multiple responses were checked by verifying the data in the questionnaire (on microfilm for students, hardcopy for new student supplements).

Finally, while many of the same items appear in both the main student and dropout questionnaires, occasionally the response codes used in the two questionnaires were different. In addition, some of the response scales used were the same as those used in earlier waves and/or HS&B but with the scale reversed. After machine

⁷ This code was used only when a critical item was missing and the retrieval oval was checked by the field interviewer, indicating that the respondent refused to answer.

editing was completed, the affected items were recoded. Student questionnaire items were recoded to match comparable items in HS&B and earlier waves of NELS:88. Then the dropout items were recoded to coincide with the student codes. Because response scales were recoded on questions that may not be strictly compatible, analysts should assess the comparability of questions when comparing NELS:88 second follow-up with earlier NELS:88 waves or HS&B. (The questionnaires that are presented in Appendix K have been modified to reflect these recodes; these questionnaires should match the data presented in the codebook that appears in Appendix J but will vary somewhat from the optical scan format instrument that was administered to NELS:88 students.)

6.2 Data File Preparation

The conventions used to assign SAS and SPSS-X variable names are as consistent as possible with HS&B and NLS-72. In those two surveys, variable names were assigned according to the survey wave and the question number. A similar system was developed for NELS:88. For example, BYS56A, is from the base year student survey, question 56, part A. Likewise, F1S7D, is from the first follow-up student survey, question 7 part D, while F2S84C is from the second follow-up student survey, question 84 part C.

Constructed variables--including statistical weights, special indicators or flags, and variables that are composites of one or more sources--are added to the files in order to promote high caliber analyses of the NELS:88 data. Certain items add information from study sources that would otherwise be unavailable to users; some items reference respondent properties to external standards that would be expensive for individual analysts to create; and other items are recodes or combinations of internal questionnaire sources. A number of composites have appeared in earlier rounds and represent a convenience for the analyst, rather than wholly new information. Some of these constructed variables will be used by nearly all users, while others will be appropriate to those seeking insights into distinctive populations, relationships or events.

Generally, the names of the base year flags, variables, and weights begin with BY; the first follow-up flags and weights begin with F1; and the second follow-up names begin with F2. If the variable is a school-level variable placed on the student file, the composite variable name begins with G8 (for grade 8 in base year), G10 (for grade 10 in the first follow-up) or G12 (for grade 12 in the second follow-up). A few composite variables that were built in the base year do not begin with the prefix "BY." These are: SEX, RACE, HISP, API, HEARIMP, HANDPAST, BIRTHMO, BIRTHYR. Over the course of the survey even basic demographics such as gender and ethnicity are re-examined and improved when and if new and/or more accurate information becomes available for particular cases (thus there is an F1SEX on the first follow-up files, an F2SEX on the second follow-up files, etc.).

The only reserved code used for all of these specially constructed variables is for missing data. For one-column variables that code is "8." Variables that are greater than one column in length are filled with "9"s (i.e., 998) in all but the right-most column. This reserved code is used when the sources for data are missing due to either item nonresponse, nonparticipation in all or part of the components of the study, or when data are missing on one or more external source files. Appendices H in the base year manual, I in the first follow-up manual and H in the second follow-up manual explain the conditions under which specific composite variables were assigned a missing code.

6.3 CD-ROM Electronic Codebook

An electronic codebook (ECB) permits PC users to interact with all of the features of a conventional hardcopy codebook and its accompanying documentation. In a very large, complex survey such as NELS:88 with multiple highly elaborated codebook text files, the Compact Disc (CD) medium provides the necessary capacity to carry a tremendous amount of data in a very compact and convenient form. CD-ROM is a form that can be copied to and read by a microcomputer. The information on CD-ROM is "Read-Only." This feature protects the data on the disk from accidental alterations, such as a user unintentionally writing over the encoded information.

In addition to numerous hardcopy codebooks that accompany magnetic tape releases on NELS:88, ECBs are also now available to users. These permit users to search for variables based on key words and names. The ECB displays question text and frequencies for each variable in order to assist users in deciding which data elements may be useful in planned analyses. The ECB is also a tool for selecting variables for subsequent analysis, writing SAS or SPSS-PC code for file construction of the designated variables, and even generating a codebook of the chosen set of variables.

More detailed information on the features of the NELS:88 ECBs and the survey waves and components for which ECBs are available appears in Chapter VII.

VII. Guide to the Data Files, Documentation and CD-ROM Electronic Codebooks

Fourteen NELS:88 study components are now available to users on magnetic tape or CD-ROM format. Magnetic tape and CD-ROM releases of the NELS:88 data contain files that are specific to one survey wave and one component, such as the second follow-up student component data. Table 7-1 displays these NELS:88 products, by study component and by survey year.

The student and dropout data sets are the central units of analysis in NELS:88. Each of the student data files may be examined as an independent entity or may be combined for observation of the maturation of the original student cohort over time. The second follow-up CD-ROM includes three waves of student component data files including:

1. **Base year data.** The base year file contains the 1988 student questionnaire data and a number of constructed variables, including the base year weight, demographic and social context measures, and cognitive test scores. There is a record in this file for every base year participant (N=24,599), regardless of whether or not the sample member was retained in the first or second follow-ups to the study.
2. **First follow-up student data.** The first follow-up student file contains a record for every student sample member in the first follow-up, regardless of actual participation and regardless of whether or not the student dropped out of school. The file includes 21 items which are common to both the first follow-up student and dropout questionnaires.⁸ This student file also contains two first follow-up student weights, first follow-up composites, including cognitive test scores and a substantial amount of new student supplementary material (demographic and other basic information collected from freshened sample members in each follow-up and from base year nonrespondents). Items collected in the new student supplement have been mapped to equivalent items that were collected in the base year. Thus, basic descriptive information is available in this section of the data file for all cases that completed either a first follow-up student questionnaire or a new student supplement.

Changes to the First Follow-Up Files. The final release of the base year, first follow-up, and second follow-up

⁸ For a complete explanation, the reader should consult section 7.2 of the *First Follow-Up: Student Component Data File User's Manual*.

Table 7-1
NELS:88 components and survey waves for which both
magnetic tape and CD-ROM products are available

	Available	Number of Variables on Public Use Version ^a
<u>Base Year</u>		
Student	Yes	410
Dropout	Not Applicable ^b	Not Applicable ^b
School	Yes	211
Teacher	Yes	238
Parent	Yes	331
Transcript	Collected in Second Follow-Up ^d	Collected in Second Follow-Up ^d
<u>First Follow-Up</u>		
Student	Yes	694
Dropout	Yes	561
School	Yes	832
Teacher	Yes	466
Parent	Not Collected ^c	Not Collected ^c
Transcript	Collected in Second Follow-Up ^d	Collected in Second Follow-Up ^d
<u>Second Follow-Up</u>		
Student	Yes	786
Dropout	Yes	577
School	Yes	385
Teacher	Yes	420
Parent	Yes	423
Transcript	Yes	--- ^d

^a The student ID number has not been included in the count of the number of variables on the public use data files. For the first follow-up school and second follow-up student files which are split into two files, the questionnaire weight has been counted only once.

^b Since by definition dropouts could only be identified and studied after the initial round of the survey, there is no base year dropout component.

^c The parent component was only conducted during the base year and second follow-up.

^d Transcripts collected during the second follow-up span the entire high school career and are available in restricted use form only. The restricted use transcript file includes 236 student-level variables and 251 course-level variables.

CD-ROM and magnetic tapes includes a number of minor changes to the cases on the file since the initial release of the first follow-up files. Four students have been verified as sample errors since delivery of the first follow-up data tape, and have been deleted from the new file released after the second follow-up surveys. Also deleted from the second follow-up data tape are 23 ineligible students who were freshened in the first follow-up. In addition, 312 of the 340 students who were Base Year Ineligible (members of the "BYI" sample) were found to be eligible for the first follow-up round in 1990, and were subsequently surveyed, while the remaining 28 were out of scope. Data for these sample members have been added to the final release of the first follow-up file. Thus, whereas the total number of students on the initial release of the first follow-up student file was 21,019 cases, the corrected final version of the public use data file instead contains 20,840 student records.

The changes in the first follow-up cases result in changes to the frequency distributions shown in the original codebooks, both hardcopy and electronic versions. Moreover, certain variables have been updated since the first follow-up file was first released.⁹ Revised weighted distributions on key analytic variables, including any altered variables, are provided with the new release so that users may review totals and percentages.

3. **Second follow-up student data.** The second follow-up student file contains a total of 21,188 records for students who are in school as well as those who have dropped out of school. It does not contain any dropout questionnaire data but does include the classification flags and other specially-constructed variables, including weights and cognitive test variables, that were built for all sample members.¹⁰ The second follow-up student data set is similar to the first follow-up file structure, with the exception of the greater number of statistical weights (cross-sectional, panel, and contextual) and cognitive test scores that are provided in this round of the survey.

The student and dropout data files released in the second follow-up of NELS:88 may be combined with data from second follow-

⁹ Among the modified variables are socioeconomic status and school control.

¹⁰ All dropout questionnaire data are provided on a separate file, accompanied by the *Second Follow-Up: Dropout Component Data File User's Manual*.

up surveys of parents, teachers and school administrators. The most powerful analyses are possible when students are viewed in the context of these fundamental influences across the four-year time frame that is now available. The NELS:88 files are designed to be merged and used to examine how differing student outcomes are related to various structural patterns, as measured by parental, teacher, and other school influences, and/or the ways in which these change over time. The contextual data files are dependent upon and subsidiary to the student files in NELS:88. **The contextual data files cannot stand alone.** The only exception is the base year school file, which is representative of eighth-grade American schools and their principals in 1988.¹¹ The first and second follow-up school components reflect characteristics of the secondary schools to which students in the contextual sample migrated after eighth grade. Since these secondary schools were not selected as a representative sample, but on the contrary appear instead as the product of student dispersion patterns, the first and second follow-up school data must be used only in conjunction with student data. Inferences from the first follow-up and second follow-up school data files cannot be legitimately made if these data are viewed in isolation from the student files.

Several types of student sample members are included in the files; therefore, the user must take care to select the correct set. Among the types of sample members in the student data set are: 1) students who were added in the first or second follow-ups to freshen the sample; 2) sample members who have participated in one, two or all three waves of the survey; and 3) Base Year Ineligible sample members who were found to be eligible and subsequently included in the first and second follow-up surveys of NELS:88.¹² Eight analytic populations, both cross-sectional and longitudinal, are now represented in the NELS:88 student sample. Different research questions apply to different student populations. In order to choose the correct NELS:88 student sample and produce accurate results, each analyst must use the proper sample identification and questionnaire availability indicators, as

¹¹ Even for the 1988 survey of parents of eighth graders and 1992 survey of parents of cohort students and dropouts--which closely resemble probability samples of parents of the relevant student populations--there are some departures from the requirements of a stand-alone sample. In particular, some unknown number of base year parents had more than one eighth grader, hence more than one chance of selection into the sample. In addition, in both 1988 and 1992, only one parent was surveyed, and that parent was self-selected.

¹² Note however that the sample of reclassified BYIs (i.e., those found to be eligible in the first follow-up and second follow-up rounds) had not been released prior to the second follow-up.

well as the correct statistical weights.

Section 7.1 introduces the reader to statistical software packages that can be used with the NELS:88 data sets and the importance of sample indicators and statistical weights in the production of accurate results. Section 7.2 includes a complete description of the content and organization of the second follow-up student data files. Finally, section 7.3 offers an explanation of the hardcopy codebook and an introduction to the electronic codebooks.

7.1 Basics for Analyses: Second Follow-Up Questionnaire and Sample Indicators, Statistical Weights and Use of Statistical Packages

The method for naming variables follows a simple pattern. "F2" refers to the second follow-up, "F1" refers to the first follow-up, and "BY" refers to the base year. An "F2" in the prefix means that the variable has been created in the second follow-up for second follow-up sample members. This is an important distinction since some variables that measure the same concept have been created for data sets in more than one round of the survey. In addition, if new information becomes available--for example, for students who have not heretofore participated in NELS:88--certain classification variables are revised to reflect this new information. The more recent the creation of a composite, the more likely that it contains the most accurate values.

7.1.1 Questionnaire/Sample Flags Included on Magnetic Tape and ECB Releases

Questionnaire Flags. One of the first steps to take in carrying out a plan for research involves selection of the proper questionnaire availability indicators. Even tentative investigations that are not statistically weighted must utilize the appropriate indicators for cases with the specified survey documents on the data file.

Eight variables have been constructed to indicate which second follow-up sample members responded to key survey documents since the base year of NELS:88. There is an indicator for each wave (1988, 1990, 1992) of the survey questionnaire, including dropout questionnaires when appropriate. There are also indicators for the presence on the second follow-up student files of cognitive test data and new student supplement data. In each of the following indicators, "1" means that the documents were completed by a second follow-up sample member and "0" means the documents were not completed. A value of "2" is present if the sample member completed a dropout questionnaire.

F2BYQFLG 1 = second follow-up sample member completed a base year student questionnaire

	0 = second follow-up sample member did not complete a base year student questionnaire
F2F1QFLG	2 = second follow-up sample member completed a first follow-up dropout questionnaire
	1 = second follow-up sample member completed a first follow-up student questionnaire
	0 = second follow-up sample member did not complete either a first follow-up student or dropout questionnaire
F2QFLG	2 = second follow-up sample member completed a second follow-up dropout questionnaire
	1 = second follow-up sample member completed a second follow-up student questionnaire
	0 = second follow-up sample member did not complete either a second follow-up student or dropout questionnaire
F2TXFLG	1 = second follow-up sample member completed a second follow-up cognitive test battery
	0 = second follow-up sample member did not complete a second follow-up cognitive test battery
	<p>This variable appears on the dropout file in order to inform users of the presence or absence of cognitive test data for these cases. The actual test scores, however, appear only on the student component data files.</p>
F2NSSFLG	1 = second follow-up new student supplement questionnaire completed
	0 = second follow-up new student supplement questionnaire not completed
F2CXTFGL	2 = second follow-up sample member is a member of the contextual sample but did not complete a second follow-up student questionnaire
	1 = second follow-up sample member is a member of the contextual sample and completed a second follow-up student questionnaire
	0 = second follow-up sample member is not a member of the contextual sample
F2TRP1FL	Indicates whether or not a sample member was a part of

both the eighth- to twelfth-grade student panel, a participant in all three rounds, and a transcript was collected for the sample member.

- 1 = The sample member is a member of the eighth- to twelfth-grade panel and transcript data is also available for the student, and the sample member completed a questionnaire in all three rounds.
- 0 = The sample member was not included in both the eighth- to twelfth-grade panel and the transcript study, or did not complete a questionnaire in all three rounds.

F2TRP2FL Indicates whether or not a sample member was a part of both the tenth- to twelfth-grade panel, a participant in the first and second follow-up, and a transcript was collected for the sample member.

- 2 = The sample member was enrolled in the tenth grade in the spring of 1990 and was member of the 1990 - 1992 panel, completed a first and second follow-up questionnaire, and transcript data is available for the sample member.
- 1 = The sample member is included in the 1990 - 1992 panel sample, completed a first follow-up and second follow-up questionnaire, and a transcript was collected for the sample member. However, the student was not enrolled in the tenth grade in the spring of 1990.
- 0 = The sample member was either 1) not a member of the 1990 - 1992 completers who also were included in the transcript component; 2) a member of the 1990 - 1992 panel and the transcript component but did not complete a questionnaire in the first follow-up or second follow-up; or 3) a member of the 1990 - 1992 panel and the transcript component but a transcript was not collected for the sample member.

Additionally, the student file includes a number of variables which indicate the availability of data from other second follow-up components for sample members. F2ADMFLG, F2TEQFLG, F2PAQFL, and F2TRSCFL indicate whether or not second follow-up school administrator, teacher, parent, and transcript data respectively are available for each sample member.

Sample/Target Population Indicators. Below are three sample indicators that have been designed for use with second follow-up statistical weights in order to generalize to appropriate respondent populations. The analyst will need to select the indicator and associated statistical weight that best suits his or

her research needs, be it a cross-sectional, longitudinal, or contextual examination of the data.

For example:

- Cross-sectional research examines students who completed a survey document in a particular survey wave of NEL8:88 (whichever is of highest interest).
- Longitudinal research requires a panel of students who completed survey documents in two or more rounds of the study (the choice of which panel to select depends upon the period of time in which events and change are best examined, given the research question).
- Contextual research investigates the effects of the family, school, and teachers that circumscribe and shape the education of students (the choice of influences selected from contextual data files depends upon the structure to be explored).

Therefore, the type of research determines which indicators should be used. The variable name for each of these sample selection indicators is as similar as possible to the variable name for the corresponding weight. The three indicators are as follows:

F2BYF1PN 1 = second follow-up sample member is in the 1988 to 1990 panel sample

0 = second follow-up sample member is not in the 1988 to 1990 panel sample

F2F1PNFL 2 = second follow-up sample member is in the 1990 to 1992 panel sample. The first-order requirement for this designation is enrollment in the tenth grade in the spring of 1990 with presence of the required survey instruments dependent upon that criterion: completion of a first follow-up student questionnaire and a second follow-up student or dropout questionnaire. Note that this is a narrower definition than merely completion of both a first follow-up and second follow-up questionnaire.

The two remaining categories differentiate among two groups of sample members who are *not* members of the sophomore panel:

1 = second follow-up sample member was not enrolled in the tenth grade in the spring of 1990 and completed a first follow-up student or dropout questionnaire and a second follow-up student or dropout questionnaire

- 0 = second follow-up sample member did not complete either a first follow-up or second follow-up questionnaire or both (regardless of enrollment status)
- F2PNLFLG** 1 = second follow-up sample member is a member of the full panel sample: 1988-1990-1992 panel (a base year student questionnaire and a first follow-up student or dropout questionnaire and a second follow-up student or dropout questionnaire were completed)
- 0 = the sample member did not complete a questionnaire in all three waves of NELS:88

Table 7.1.1-1 provides a summary of indicator values to be used when selecting populations for analyses.

Weights. The NELS:88 data files are designed to be used as weighted data sets in all analyses. Due to the complexity of the NELS:88 sample design, estimation and inference will most likely be inaccurate if the data are analyzed on an unweighted basis. Clustering, multistage selection, and disproportionate sampling all contribute potential bias and various degrees of unreliability, which can only be avoided by using the weights provided to analyze specific subsets of the sample.

In the variable name for statistical weights, the suffix "WT" is used to distinguish these from the special sample indicators described above. When the user combines a sample indicator with the appropriate weight, population estimates are produced. Seven second follow-up statistical weights have been created for the second follow-up student and dropout data files. Although four panel weights (F2F1PNWT, F2PNLWT, F2TRP1WT, and F2TRP2WT) are listed, no contextual panel weights have been constructed.

- F2QWT** use for producing weighted twelfth-grade student statistics in cross-sectional analyses.
- F2F1PNWT** use for producing weighted student panel statistics when both first follow-up and second follow-up data are employed in the analyses.
- F2PNLWT** use for producing weighted student panel statistics when all three survey waves (base year, first follow-up and second follow-up) data are included in the analyses.
- F2TRSCWT** use for conducting cross-sectional analysis of transcript data.

Table 7.1.1-1 Summary of NELS:88 populations, samples, level of analyses, sample identification flags, and weights

<u>Population of Interest</u>	<u>Sample and Sample N</u> (Participants)	<u>Level of Analysis</u>	<u>Sample ID Flag</u>	<u>Weight</u>
1988 8th Graders in 1990 (Panel): The population of 1988 eighth graders two years later (as of 1990).	Base year retained sample members who completed both a base year and first follow-up questionnaire. N=17,424. Note: undercoverage bias; 5% of potential base year sample excluded.	Cross-wave, longitudinal level of analysis.	Use the first follow-up student data file and select for the 8th to 10th grade panel, F1PANFLG=1.	Use the panel weight (F1PNLWT).
1988 8th-Grade Cross-Section: The population of all students enrolled in the eighth grade in 1988.	Base year selected sample members who participated in the base year. N=24,599. Note: undercoverage bias; 5% of potential base year sample excluded.	Cross-sectional level of analysis.	Use the base year student data file and select for BYQFLG=1.	Use the base year questionnaire weight (BYQWT).

Table 7.1.1-1 (cont.) Summary of NELS:88 populations, samples, level of analyses, sample identification flags, and weights

<u>Population of Interest</u>	<u>Sample and Sample N</u> (Participants)	<u>Level of Analysis</u>	<u>Sample ID Flag</u>	<u>Weight</u>
<p>1990 10th-Grade Cross-Section: The population of all students enrolled in the tenth grade in 1990.</p>	<p>Representative sample of students enrolled in tenth grade in the spring term of 1990. Includes freshened students and excludes dropouts and out-of-sequence sample members. N=17,544 on the original F1 release. On the F2 re-release, BYI's who completed a F1 questionnaire appear producing a new N=17,754.</p>	<p>Cross-sectional analysis; Trend analyses with HS&B 1980 sophomores and F1 1990 sophomores.</p>	<p>Use the first follow-up student file and select for F1SEQFLG=0.</p>	<p>Use the first follow-up questionnaire weight (F1QWT).</p>
<p>1990 (First Follow-Up) Cross-Section: The population of all first follow-up eligible persons who were either in the eighth grade during the 1987-1988 school year or were in the tenth grade during the 1989-1990 school year.</p>	<p>All first follow-up 1990 sample members; combines 1988-eligible eighth-grade cohort and 1990 tenth-grade freshened sample. N=19,264 on the original release. On the F2 re-release, BYIs who completed a F1 questionnaire appear producing a new N=19,394.</p>		<p>Use the first follow-up student file and select for F1STAT=0.</p>	<p>Use the first follow-up questionnaire weight (F1QWT).</p>

Table 7.1.1-1 (cont.) Summary of NELS:88 populations, samples, level of analyses, sample identification flags, and weights

<u>Population of Interest</u>	<u>Sample and Sample N</u> (Participants)	<u>Level of Analysis</u>	<u>Sample ID Flag</u>	<u>Weight</u>
1988 8th Graders in 1992 (Panel): The population of 1988 eighth graders four years later (as of 1992).	Base year retained sample members who completed a questionnaire in all three waves of NELS:88--base year, first follow-up, and second follow-up. N=16,489. Note: undercoverage bias; 5% of potential base year sample excluded.	Cross wave, longitudinal level of analysis.	Select for F2PNLFLG=1.	Use the second follow-up panel weight (F2PNLWT).
1990 10th Graders in 1992 (Panel): The population of 1990 sophomores two years later (as of 1992).	Representative sample of students enrolled in tenth grade in the spring of 1990 (see definition of tenth grade cross-section above) who completed a questionnaire in both the first follow-up and second follow-up. N=16,749.	Cross wave, longitudinal level of analysis; longitudinal trend analyses with HS&B 1980 sophomore cohort and F1 1990 sophomore cohort.	Select for F2F1PNFL=2.	Use the second follow-up panel weight (F2F1PNWT).

Table 7.1.1-1 (cont.) Summary of NELS:88 populations, samples, level of analyses, sample identification flags, and weights

<u>Population of Interest</u>	<u>Sample and Sample N</u> (Participants)	<u>Level of Analysis</u>	<u>Sample ID Flag</u>	<u>Weight</u>
1992 12th-Grade Cross-Section: The population of all students enrolled in twelfth grade in 1992.	Representative sample of students enrolled in twelfth grade in the spring of 1992. N=16,114. Includes freshened students and excludes dropouts and out-of-sequence sample members.	Cross sectional analysis; Trend analyses with NLS 1972 seniors, HS&B 1980 seniors, and NELS:88 1992 seniors.	Select for F2SEQFLG=0.	Use the second follow-up questionnaire weight (F2QWT).
1992 (Second Follow-Up) Cross-Section: The population of all second follow-up eligible persons who were in eighth grade during the spring of 1988 or in tenth grade during the spring of 1990 or in twelfth grade during the spring of 1992.	All second follow-up 1992 sample members; combines 1988-eligible eighth-grade cohort and 1990-eligible tenth grade cohort and 1992 twelfth-grade freshened sample. N=19,220.		Select for F2STAT=00.	Use the second follow-up questionnaire weight (F2QWT).

Note: For information on how to use school, parent and transcript data--and respective weights--with student data, consult the separate data file user's manuals for each of these components.

- F2TRP1WT** use for conducting panel analyses using the transcript component data with test and questionnaire data for the panel of 1988 eighth graders four years later (1992).
- F2TRP2WT** use for conducting panel analyses using the transcript component data with test and questionnaire data for the panel of 1990 tenth graders two years later (1992).
- F2CXTWT** use for producing weighted student contextual component statistics, in conjunction with cross-sectional analyses that also involve school administrator and/or teacher data.¹³

Cross-sectional analysis of second follow-up student data requires that the F2QWT weight variable be applied. Longitudinal analyses, on the other hand, require use of F2F1PNWT or F2PNLWT panel weights, with the difference hinging upon the time points that define the student panel that the user wishes to examine. Thus, if F2PNLFLG is used to select cases from all three waves of NELS:88, then F2PNLWT is the correct statistical weight in the analyses. Similarly, if F2QFLG is used to select second follow-up student respondents, then F2QWT should be used.

Full Sample Member/Population Indicators. The variables described in this section apply to the full student sample, as opposed to specific portions limited by participation, enrollment status or eligibility. G8COHORT indicates whether or not the student is a member of the eighth-grade cohort; that is, whether or not the student was enrolled in the eighth grade during the 1987-88 school year and therefore eligible to complete a NELS:88 base year student questionnaire. Similarly, G10COHRT indicates whether or not the student is a member of the tenth-grade cohort (enrolled in the tenth grade during the 1989-90 school year and eligible to complete a first follow-up student questionnaire). G12COHRT indicates whether or not the student is a member of the twelfth-grade cohort (enrolled in the twelfth grade during the 1991-92 school year).

¹³ Panel analyses that use school administrator and/or teacher data from the NELS:88 base year or first follow-up in conjunction with second follow-up data should apply the F2CXTWT with caution. Because of factors such as nonresponse in the base year and first follow-up, this weight is not as precise as a contextual panel weight would be, and analysts should assess their results for bias. Results should also be compared with those obtained by utilizing alternative weighting "approximations," e.g., F2TRP1WT for the 1988-1990-1992 panel.

Second Follow-Up Status and Participation Flags. F2STAT indicates the final disposition in the second follow-up for each sample member on the file. Categories include participation, unlocatable, student or parent refusal, ineligibility due to a mental or physical disability or to a language barrier, residence outside the U.S., or deceased.

F2DOSTAT indicates enrollment status, either dropout or student, as of the second follow-up only. This indicator also permits identification of dropouts according to the "first follow-up definition" of a dropout (i.e., sample members receiving no formal instruction) and the "second follow-up/HS&B definition" of a dropout (i.e., combines sample members who are enrolled in an alternative or non-traditional high school equivalency program with those receiving no formal instruction [who completed the dropout questionnaire] as well as sample members who have earned a GED or other alternative credential [who, as school equivalency completers, were administered the student questionnaire]).

F2SEQFLG indicates whether or not participating sample members are currently in grade sequence; that is, enrolled in the twelfth grade. This variable also identifies early graduates, dropouts, alternative completers, and other out-of-sequence sample members, regardless of their participation status.

F2SMPFLG indicates how and when sample members were brought into the study. Valid categories include eighth-grade cohort member, tenth- or twelfth-grade freshened student, or base year ineligible student.

Several other indicators provide additional information about sample members such as whether or not the sample member was an early graduate, a dropout, ineligible or out-of-scope, or freshened into the sample. These variables are: F2EGDFLG, F2F1DOST, F2EVDOST, and F2TRSTYP. Because school records may contradict other sources of sample members' enrollment status, the NELS:88 student and transcript component files include F2TRSTYP which identifies inconsistencies among different sources of a sample member's enrollment status, including F2DOSTAT and F2RTROUT on the transcript data file. See Appendix H for a complete description of F2TRSTYP.

Universe Variables. As in every longitudinal survey, the complexity of NELS:88 has increased with each successive survey wave. The changing numbers of cases delivered in each round may be one of several perplexing anomalies to users. The "universe variables" are designed to explain how the status of sample members has changed from one wave to another. The first of these, F2UNIV1, is a set of over one hundred mutually-exclusive categories that was designed to encompass each and every sample member ever in the study. It describes how and when the sample member entered NELS:88 and the situation of the sample member in the base year, first follow-up, and second follow-up. Abbreviations for the SAS and

SPSS-X value label cards provide this information in the character lengths allowed by those programs. These abbreviations are:

BY = Base Year
F1 = First Follow-up
F2 = Second Follow-up
I = Ineligible for questionnaire administration
(mental/physical disability, language barrier)
A = In-school, in-grade
B = In-school, out-of-grade
DO = Dropout
E = Eligible for questionnaire administration
FR = Freshened
NA = Not Applicable (status description for rounds prior to
that in which one was freshened into sample)

X = Out-of-scope (deceased, out-of-USA)
? = Status unknown

Note that a status is attached to each round (BY, F1, F2) in each valid category of this value. Examination of the categories (see the codebook distributions, for example) reveals that situations did change over time. For example, base year ineligible students were subsequently re-surveyed and some were discovered to be capable of completing the survey in the first and/or second follow-ups. Other sample members moved out of the country in a later round and were defined as "out-of-scope" for that round (note that some had returned to the U.S. by the second follow-up and were then once again in-scope for NELS:88). Similarly, freshened students at the secondary school stage did not participate in the base year.

Four additional universe variables are provided, each with a more limited descriptive mission than F2UNIV1. These variables account separately for the information that is combined in the first universe variable. F2UNIV2A reports how sample members initially entered NELS:88. Categories are base year eligible, base year ineligible, or freshened (in either the first or second follow-ups). F2UNIV2B reports the base year status of all sample members: freshened in either the first or second follow-ups (and thus not at that time an active sample member), in school and in the appropriate grade, ineligible in that round due to a mental, physical or a linguistic barrier. F2UNIV2C reports the first follow-up status of sample members. Categories are freshened in the second follow-up of NELS:88, in school and in the appropriate grade, in school but not in the expected grade for the cohort, dropout, ineligible for this wave, out-of-scope (deceased or not in the U.S. during this round of the study), or status currently unknown. Finally, F2UNIV2D reports on the second follow-up status of each sample member. Valid possibilities are in school and in the expected grade, in school but not in the expected grade, dropout, ineligible, out-of-scope or status unknown in this round of NELS:88.

7.1.2 Packaged Statistical Programs

The procedures recommended for analyses of NELS:88 data with the SAS or SPSS-X are outlined in Appendix I. Both the magnetic data tape and the ECB on CD-ROM can generate the appropriate control cards for each of these statistical packages. It is also possible for analysts to create an SPSS-X system file from a SAS system file (or vice-versa).

7.2 Content and Organization of the Data Files

The second follow-up student data file contains a record for 21,188 participating and nonparticipating sample members. Of these, 16,842 participated as students, 2,378 participated as dropouts/dropouts in alternative programs, and 1,968 did not participate in this round of the study. Note that data elements from the dropout questionnaire are not included on this file but appear instead on the second follow-up dropout files in a separate release.¹⁴ The 21,188 records include 13 students who were freshened in the second follow-up but were temporarily out of scope (out of the country). While these records are included in the file, they do not have any weights, and no questionnaire data are associated with these cases. Additionally, the 21,188 includes 235 sample members who did not complete a questionnaire in the base year, first follow-up and second follow-up. These sample members who were nonresponders in all three waves do not have weights but have been included on the tape for analysis of unit nonresponse.

The raw data file contains 796 questionnaire variables on the 17,192 sample members who completed a student questionnaire. In addition, a large number of special flags, composite variables, and three statistical weights are included on the second follow-up student releases. The student component file also includes school-level composite variables that have been constructed for dropout sample members based on the last school attended by the dropout.

The record layout for the second follow-up appears in Appendix G. The layout shows in detail the organization of the variables within each record on the file. The variables are grouped into similar logical sets as discussed below. Each data item is referred to by its SAS (SPSS-X) variable name, as defined in the control cards provided with the data file.

Twelve files related to the student surveys, four for each wave of NELS:88, have been released. They are:

1. **The base year** raw data file with the following items for each sample member participating in the base year:

¹⁴ See the *NELS:88 Second Follow-up: Dropout Component Data File User's Manual* for complete documentation of the dropout files.

- a. Randomized ID number (positions 1-7),
 - b. Base year student questionnaire data (positions 8-358),
 - c. Base year weight, flags, and composites (positions 359-548);
2. SPSS-X control cards for the base year file;
 3. SAS control cards for the base year file;
 4. SAS system file for the base year data;
 5. **The first follow-up** raw data file consists of the following items:
 - a. Randomized ID number (positions 1-7),
 - b. First follow-up student questionnaire data, along with a subset of dropout questionnaire data that matches student questions (positions 8-664),
 - c. First follow-up weights, flags, and composites (positions 665-857),
 - d. First follow-up new student supplement data with equivalent base year data mapped into the new student supplement items (positions 858-932);
 6. SPSS-X control cards for the first follow-up student file;
 7. SAS control cards for the first follow-up student file;
 8. SAS system file for the first follow-up student file;
 9. **The second follow-up** raw data file consists of the following items:¹⁵
 - a. Randomized ID number (positions 1-7, file 1),
 - b. Second follow-up student questionnaire data (positions 8-653, file 1),
 - c. Second follow-up weights, flags, and test scores; and other composites (positions 8-268; 339-603, file 2),
 - d. Second follow-up new student supplement data with equivalent base year data mapped into the new student supplement items (positions 269-338, file 2). These variables are described more fully in section 7.2.3 below.

¹⁵ Because the number of variables in the student data file exceeds the amount allowed by SPSS-X, the data file has been split into two files.

10. SPSS-X control cards for the second follow-up student file;
11. SAS control cards for the second follow-up student file; and,
12. SAS system file for the second follow-up student file.

7.2.1 Identification Codes

The first variable on all of the raw data files, STU_ID, is a unique seven-digit student identification code. This number remains with the student throughout NELLS:88. To link student records across two or more waves of the survey (1988, 1990, and 1992) or between survey components (student, dropout, teacher, school, parent, and transcript), analysts should use STU_ID.

The student ID code consists of a five-digit base year school ID followed by a two-digit student code. Though both sets of numbers were randomly assigned to maintain confidentiality, the IDs contain embedded linking, stratum, and PSU information.¹⁶ Students added to the first or second follow-ups through freshening were linked to a core sample member. The base year school ID of the linked student was used as the root of the added student's ID. Thus, in all cases, the student ID links the students and dropouts to a base year school.

7.2.2 The Student Survey Instruments

The logical record length, block size, and record layout for the second follow-up student data file appear in Appendix G. This layout shows how variables are ordered within each student data record on the file. Information was collected directly from respondents in two source documents, the student questionnaire, and the new student supplement (NSS).¹⁷ Results from the student questionnaires appear at the beginning of each student data record, in the same order as the questions appear in the printed

¹⁶ Analysts who are employing variance estimation software should note that the student ID reflects the NELLS:88 sampling plan in the following way: the left-most two digits of the ID represent the stratum identification number for the case; the middle three digits are the primary sampling unit (PSU) for the school; and the last two digits identify the student uniquely within the stratum and PSU.

¹⁷ Information collected in the dropout questionnaire is reported on a separately-released data file. Data collected for early graduates are included on a separate CD-ROM that is not in the ECB format. See section 7.3.2 for additional information about this CD-ROM.

questionnaires; data collected in the NSS appear at the end of each student data record. Appendix K contains a copy of the second follow-up student questionnaire and the new student supplement.

The variables in the record layouts are identified by the SAS and SPSS-X variable names that have been designated for each data element. No more than eight characters may comprise a SAS or SPSS-X variable name. The first three characters of the variable names from the student questionnaire or the NSS indicate the survey wave in which the variable was created, as well as the source document used to collect the information. Thus, `BYS` in the prefix of the variable name indicates a base year student questionnaire item. `F1S` or `F2S` in the prefix of the name refers to a student item in either the first follow-up (the 1990 round) or the second follow-up (the 1992 round). `F1N` or `F2N` refer to the NSS source document for either the first follow-up or the second follow-up. The naming scheme for items that report student responses is completed by the suffix of the variable name, which consists of the question number and part. For example, `F1S23H` is question 23, part H from the first follow-up student questionnaire; `F2N19E` is question 19, part E from the new student supplement.

The New Student Supplement. In the first and second follow-ups, new students were added, or freshened, to create nationally-representative sophomore and senior cohorts. It was necessary to collect basic social information, already asked of earlier participants, for the new students. It was also essential to collect the same type of information for nonresponders in early rounds who subsequently participated in the study. Newly eligible students, freshened students, and former nonresponders were therefore asked to complete new student supplement booklets.

The NSS collects some 70 respondent-reported facts about new participants, their families, and households. The characteristics reported in the NSS are routinely used as key independent or control variables in analyses of education outcomes. For this reason, the responses of new participants are merged with responses of earlier participants to the same questions. These earlier responses have already been released in the original data sets in which they were collected. The NSS variables are constructed by merging the same characteristics for all cases, regardless of the year or the source document in which the characteristics were first collected from the student. The NSS variables were constructed in this way as a convenience to users, who would otherwise need to locate and merge the information with earlier waves on their own. Note that it is a simple matter to divide these items by subgroups of students who participated in each survey wave through the use of the flags described in section 7.1.1. For example, `F2NSSFLG` allows users to select only cases on which these data were collected in the second follow-up survey wave and review their response distributions separately.

The NSS begins with information on the student's mother and

father, such as education level attained, employment, and occupational category. These are important indicators of the level of social resources that may be available to students. They are also fundamental assets that mediate the efforts of teachers, schools, and communities to educate students.

The NSS collects the number of brothers and sisters present in the family, and the student's birth order within it, both of which are meaningful to analysts of personality formation and behavior. Whether or not siblings have left school prior to graduation appears to be a key predictor, when combined with a number of other predictors that are measured within NELS:88, of the probability that a sample member will also subsequently drop out of school. Religious background, reading and writing fluency in the student's native language, and whether or not the student first learned to speak in English or another language are additional patterns that may influence how learning occurs.

The supplement captures information that reveals the student's exposure to a variety of material goods in the home which may affect education. For example, access to personal computers, books in the home, and specific places for study are important elements of the household and family context that affect the educational development of the student.

Many NSS items become building blocks in the construction of composite variables, such as socioeconomic status, which combines items from both student and parent source documents. Appendix H lists the specifications for socioeconomic status and other composites provided on the second follow-up data records.

The new student supplements are subjective reports by student respondents. The distribution of race and gender in particular may differ in these self-reports from the composite variables F2RACE1 and F2SEX that appear elsewhere on the student data records. The composites replace missing values in the NSS self-reported characteristics with information from school records or the results of imputation procedures (i.e., inference of gender from first name). Additionally, in one particular instance with high student over-reporting--American Indian race/ethnic status in the base year--parent reports were used to override student reports when the two disagreed.

F2N17 (student race in the NSS) is a qualitatively different measure from F2RACE1. The first variable reports the sample member's own identification with a racial or ethnic category at a particular moment. Many researchers are keenly interested in how students respond to the stimulus, "Which best describes you?" when presented with a set of categories. Even refusals to this item are of interest to methodologists, questionnaire designers, and scholars who examine race relations in America. It is desirable to preserve the original responses for such investigators. Other analysts, interested in race/ethnicity mainly as a control

variable, will prefer the composite that uses additional information sources (and may give preference to one source over another if they disagree) to supplement student self-reports.

In order to clarify these differences for users, the codebook entries for the race and gender NSS items include comments on the meaning of these student reports and point users to the race and gender composite reports as well.

7.2.3 Composite Variables

Composites variables are constructed in order to enhance substantive analyses. Since research questions frequently require independent or control variables such as the urbanicity of the school, the socioeconomic status of the family, or the gender of the individual, a large set of classification variables has been carefully constructed and added to the records. Complete specifications used to create these composite variables can be found in Appendix H for the second follow-up. (In the respective data file user's manuals, see also Appendix H for base year composites and Appendix I for first follow-up composites.)

Most composite variables are constructed from two or more sources. These may be combinations of questionnaire items from the same or different NELS:88 data files, in the same survey year or across different survey waves. Some composites are drawn from an external sampling resource that is unavailable to users, or utilize an external conceptual scheme in order to rank order or otherwise recode survey data. A few composites are sufficiently central to analyses that they are constructed in each round of the survey. Some values should change over time; for example, if a student transfers from one school to another, then school control type, urbanicity, region and so on may change as well. Some variables, such as race/ethnicity and gender, should in theory be constant for an individual student over time, yet in practice may change if new information improves upon the old. For example, a race/ethnicity composite is constructed for all student sample members, regardless of actual participation in NELS:88. In the situation in which a former nonparticipant later takes part in the survey, the value of the race composite may in very rare instances change from a value that had been imputed on earlier data sets. Such differences over time illustrate how the validity of certain classification variables is strengthened over time. The most recent round in which such a variable appears contains the best information for students who participated in that wave of NELS:88. The derived variables are described in detail in Appendix H along with flags and weights in the order in which they appear on the data file.

Demographic Composites. Many of the NELS:88 composite variables are respondent demographic characteristics. F2SEX represents student gender while F2RACE1 and F2RACE2 represent the first of two race/ethnicity composites that have been constructed for the second follow-up. These variables are important to so many

research questions that missing data cannot be tolerated. Note that this is a different approach than the methods described for the new student supplement variables, upon which several composites build and in which missing values are retained. The values for each of these characteristics were taken directly from the analogous first follow-up composites or from second follow-up new student supplements. If these sources were not available or contained missing data, sample member gender was taken from base year school rosters. Any cases that still suffered from missing values had gender imputed from the respondent's first name or if that could not be done unambiguously, the value for F2SEX was randomly assigned. F2RACE1 also was constructed from several sources of information. The first source was the student self report (from either the base year student questionnaire or the first or second follow-up new student supplement). If the student information was missing or, for student-reported race of American Indian, inconsistent with that of the base year parent report, the values from the parent questionnaire were used. If F2RACE1 was still missing, the race as identified on the school roster was used. For sample members who report in F2RACE1 that their race is Asian, Pacific Islander or Hispanic, F2API and F2HISP divide these categories into subcategories. F2BIRTHM and F2BIRTHY indicate the month and year in which each sample member was born.

Socioeconomic Status. The second follow-up files contain three versions of a continuous variable, "F2SES-", which indicates the sample member's socioeconomic status. F2SES1 was derived from the base year parent questionnaire data, the base year student questionnaire data, or the first or second follow-up new student supplement data. Both F2SES2 and F2SES3 are constructed with second follow-up parent questionnaire data. F2SES3 incorporates the 1989 revision¹⁸ of Duncan's Socioeconomic Index (SEI), whereas F2SES1 and F2SES2 utilize the original (1961)¹⁹ version that was used in NLS-72, HS&B, and the NELS:88 base year and first follow-up.²⁰ F2SES1 has been constructed for all sample members and appears on the student and dropout files, but F2SES2 and F2SES3 appear only on the parent component data file and, therefore, have only been constructed for the subset of student and dropout sample

¹⁸ Nakao, K., and Treas, J. (1992). *The 1989 Socioeconomic Index of Occupations: Construction from the 1989 Occupational Prestige Scores: General Social Survey Methodological Report No. 74*. Chicago: NORC.

¹⁹ Duncan, O.D. (1961). "A Socioeconomic Index for All Occupations." In *Occupations and Social Status*, A.J. Reiss et al. eds, New York: Free Press.

²⁰ Note that one value in the occupational prestige scale was transposed in earlier releases of the socioeconomic status composite variable and has been corrected for all second follow-up releases of F2SES1.

members for whom parent data were collected.

F2SES1 is constructed in the same way as the first follow-up socioeconomic status measure. Student socioeconomic status is estimated from the base year parent questionnaire, the primary source for F2SES1; only F2SES2 and F2SES3 utilize second follow-up parent data. The composite is constructed with the values of five standardized components: father's and mother's educational levels, father's and mother's occupations, and family income. For cases without parent data, student data were used from either the base year student questionnaire or the first or second follow-up new student supplement. The first four components from the student data are the same as the components used from the parent data and a ranking of material possessions was substituted for family income. F2SES1Q is simply the F2SES1 quartile to which a respondent belongs.

Composites of School-level Characteristics. The composites of school-level characteristics provide information on key characteristics of sample members' second follow-up school. Note that school-level composites for students who have dropped out of school appear only on the second follow-up student component data files. Special coding of verbatim responses in the dropout questionnaires was conducted to capture the "school last attended" by that the sample member reported in that document. School-level composites were then built for dropouts using the "school last attended."

G12CTRL1 classifies the student's second follow-up school by type of control: public, Catholic or other private, with private schools divided into other religious, no religious affiliation, or affiliation unknown. G12CTRL1 is primarily reported from the school administrator questionnaire.

G12URBN3 is a three-category composite that reflects the type of place in which the student's public school district, Catholic diocese, or, for other private schools, county is located. The categories are urban, suburban and rural. The information was obtained from QED, or when missing, looked up in the U.S. Bureau of the Census, *Statistical Abstract of the United States: 1992* (112th edition), Washington DC, 1992, pages 896-904, and added to the files. This composite is analogous to the variable that was used in HS&B and in NELS:88 sampling.

G12REGON indicates in which of the four U.S. Census regions the student's second follow-up school is located (Northeast, Midwest, South, West). It is created by collapsing the values of the state in the school address.

On the restricted use student file, G12STATE identifies the state in which each sample member's school was located.

Composites of Student-level Characteristics. Two composites

capture the student sample member's perception of how chance, versus one's own actions, affects the way that life unfolds. Both draw on items that are in the second follow-up student and dropout questionnaires. The first "Locus of Control" construct is designed to be as comparable as possible with similar variables that were created for the NLS-72 and HS&B surveys. In order to achieve this comparability, F2LOCUS1 comprises fewer items than F2LOCUS2. A standardized quartile (created by applying the student weight to scores) is represented in F2LOCU2Q.

Similarly, two constructs that attempt to capture the student sample member's level of self-esteem appear on the parent files. The first, F2CNCPT1, is designed to be as comparable as possible to parallel items in NLS-72 and HS&B, while the second, F2CNCPT2, uses the full set of second follow-up student questionnaire items that are available for operationalizing the concept. The quartile, F2CNCPT2Q, is a weighted reflection of the more elaborated self concept measure.

F2F1SCFL indicates whether or not a sample member's school data were collected from the same school during the first follow-up and second follow-up. This variable does not indicate that a student was at the same school continuously (some small portion of students may have moved from a first follow-up school, then subsequently returned to the school by the time of data collection in the second follow-up). This variable is included on the restricted use student file.

Transcript Flags, Indicators, and Composites. A large number of transcript variables are included on the student file and are described in detail in Appendix H. These variables include transcript flags, school-level composites, student and course-level composites, New Basics composites, NAEP-equivalent New Basics composites, and subject-area composites.

Cognitive Test Variables. Cognitive test variables are also included on the student file and are described in detail in Appendix H.

7.3 Guide to the NELS:88 Codebooks

The codebooks that have been provided for each wave of the survey fully describe and assist with interpretation of each of the variables on each of the data files. The codebooks summarize all key information for each data element, including:

- the variable name, question number and content;
- the tape position and format on the file for each variable;
- valid and/or missing responses to each item; and,

- the unweighted frequency counts, percents, and weighted percents for each response category.

Two related types of codebooks are provided for NELS:88--a hardcopy codebook and an electronic codebook (ECB). Both forms of the codebook chronicle the details analysts need to interpret properly the results of each item: the exact wording of the question that was presented to the respondent, the distribution of all legitimate answers among survey participants, the location and type of data element for each variable on the file, as well as names and labels provided for use with statistical software. For some items the basic presentation is supplemented with additional notes about using the data. The first type of codebook, the hardcopy codebook, is central to the documentation that is included in the NELS:88 data user manuals. Hardcopy codebook displays are described and illustrated in section 7.3.1 below.

The second type of codebook is the NELS:88 electronic codebook (ECB). The electronic print files that are produced by the hardcopy codebook software are used as the foundation (the input files) for the ECB software. ECBs provide several advantages. First, the NELS:88 ECBs reside on CD-ROM and, given the right equipment and software, can be accessed by and copied to a user's own personal computer. The NELS:88 data sets have also been released on CD-ROMs, a far more concentrated medium for archiving information than magnetic tapes. The PC mode is both more convenient and less expensive than mainframe operations for most users. Second, ECBs permit users to scroll through the same variables and survey results found in all versions of the codebooks electronically. In addition, analysts interact with the ECB software to select only those data elements needed for the user's specific analyses. The result is a user-controlled subset of the variables that is fully equipped with the tools required for statistical analysis. The labor-intensive steps that were formerly required to accomplish these preliminary steps to analysis, such as typing in exact variable names, have been rendered obsolete by the ECB system. Additional information on ECBs is given in section 7.3.2.

7.3.1 Hardcopy Codebooks in NELS:88 Data User's Manuals

Both the hardcopy and the ECB versions of the NELS:88 codebooks contain the basic information available on each variable in the NELS:88 data sets. Therefore, even those readers who plan to use ECBs should be familiar with the material in this subsection in order to take full advantage of the ECB.

Figure 7-1 is an illustration of the information provided in the codebooks for each data element. Each portion of this example is numbered and explained below.

Figure 7-1
An entry in the student codebook

(1) Question 7G (2) Tape Pos. 20-20
(3) Format: I1

(4) F2S7G (5) FIGHTS OCCUR OFTEN BTWN RACIAL/ETHNIC GROUPS

(6) Fights occur often between different racial or ethnic groups

(7) <u>RESPONSE</u>	(8) <u>CODES</u>	(9) <u>FREQ</u>	(10) <u>PER-CENT</u>	(11) <u>WGTD PCT</u>
STRONGLY AGREE . . .	1	894	4.2%	5.4%
AGREE	2	3019	14.2%	18.2%
DISAGREE	3	7736	36.5%	47.0%
STRONGLY DISAGREE .	4	5223	24.7%	29.4%
 (12) RESERVED CODES: NONRESPONDENTS & DROPOUTS				
		3996	18.9%	(MISS)
MISSING.	8	320	1.5%	(MISS)
Totals:		21188	100.0%	100.0%

Explanations:

1. Question number: In the student files, question number is the same as the student questionnaire item number for variables taken directly from the student questionnaires. Composite variables and other items such as flags and weights have variable names that reflect their content.
2. Tape position: This item gives the starting and ending tape position of each variable on the data tape.
3. Variable format: This item indicates the type of variable, its width, and the number of positions following the implicit decimal point, if any.
4. SAS and SPSS-X variable name: Each variable in the data set is identified by a unique SAS and SPSS-X variable name. Data indicators (such as flags and status codes) and composite variables are given mnemonics that help identify them. For example, G12REGON is used for Grade 12 Census region, and F2SES1 is used for one of three second follow-up socioeconomic status variables. Users should refer to the variable by its SAS (SPSS-X) variable name in any computing procedures, rather than by its question number.

Figure 7-1 (continued)
An entry in the student codebook

5. SAS (SPSS-X) variable label: A short variable label appears after the variable name. This label is the same as that which appears on the SAS (SPSS-X) data definition cards included on the tapes or CD-ROM.
6. Original question wording: This reproduces the exact question wording as it appeared in the questionnaire.
7. Response categories: This item provides either the original response categories in the case of questionnaire items or the recoded or constructed response categories for special variables such as a statistical weight. For display in the codebooks, continuous or very sensitive variables have been recoded to collapse all valid values into one or a few response categories. This allows the codebook tables to show the frequency counts, unweighted percentages, and adjusted weighted percentages for continuous variables without printing each distinct value that the variable can take. These value labels are not the same as those on the SAS (SPSS-X) data definition cards. Condensed value labels that do not cause truncation problems are provided with the data definition cards.
8. Response codes: This item provides the actual numerical codes that appear on the data tape in the tape position specified (except for continuous variables, where the actual values that appear on the tape have been recoded to produce the frequency counts and percentages). Certain codes, discussed below, are reserved to indicate missing data, legitimate skips and so forth.
9. Frequency counts: This item shows the unweighted frequency counts for all records that were processed, including records that have missing data codes, legitimate skips, and so forth.
10. Unweighted percentage frequencies: This column displays the frequency counts of item 7G as percentages. All records that were processed are included.
11. Weighted percentage frequencies: This column displays percentages based on response counts weighted up to the relevant population. Cases with reserved code values are excluded from the computation.

Figure 7-1 (continued)
An entry in the student codebook

12. Reserved codes: In this data set certain codes, termed "reserved codes" have been chosen to always stand for certain situations. These reserved codes and their interpretations are:

6=multiple response more than one response where only one response was called for.

7=refusal respondent refused to answer an item or refused to resolve a multiple response where only one was called for, either at the time of the questionnaire administration or at telephone follow-up;

8=missing data data that should be present for this respondent is missing, but respondent did not necessarily refuse to provide data;

9=legitimate skip because of responses to preceding questions, data for this item should not be present for this respondent; that is, the value is legitimately missing.

These reserved codes are the same as those used in the NLS-72 and HS&B surveys. The codes as listed above apply to variables with single-column data fields. For variables with fields greater than one column, the left-most columns are filled with 9's (e.g., 96, 996, 9996).

Note that in the example shown in Figure 7-1, cases representing nonrespondents and dropouts (the latter do not respond to the student questionnaire) are shown on a separate line and represent nearly 19 percent of the total distribution.

Finally, additional comments and notes may be included and displayed below the standard information in the codebooks described in Figure 7-1. These comments alert researchers to the potential for nonresponse bias, a relation to another similar variable or composite, a recoding of a continuous variable in order to improve the codebook presentation, or to recodes or suppressions of sensitive data for confidentiality purposes.

7.3.2 The NELS:88 Electronic Codebook System (ECB)

The electronic codebook combines the convenience, simplicity and cost efficiencies of personal computers (PCs) with CD-ROM technology. Thousands of NELS:88 variables, the extensive statistical software programs and commands that transform the data in analyses, and electronic versions of data user manuals reside on a single CD-ROM. All are accessible with the MS-DOS operating system and statistical and word processing software that the user is likely already accustomed to working with on his or her own PC; however, a user must already have access to PC-SAS or SPSS-PC. Virtually all steps that must be undertaken prior to actual analysis on the data files may now be conducted within the ECB.

The ECB software is designed to acquaint the user with the available survey measures and responses by means of on-line, fully documented codebooks. Users may browse through the documentation, searching on both variables names, labels, and question text to find items that are suitable for the research question at hand. The final version of the ECB includes weighted and unweighted frequency distributions. Users can move quickly in the ECB between questionnaire items, sample indicators, composite variables, or between components of the study and may select variables of interest, up to 255 variables per session. A window shows how many variables have been tagged at any one time. The process culls a set of variables, and only those variables, that are appropriate to the user's own research issue. Since variable names and labels are already in electronic form on the ECB, onerous tasks (such as typing in this information) that were formerly necessary are eliminated. The ECB permits users to write SAS-PC or SPSS-PC program code and/or command statements in order to construct system files of the selected variables. Finally, a print file of a codebook containing the frequencies for only the tagged items is another ECB option. The print file may subsequently be used to generate individualized hardcopy codebooks of the selected variables, providing a convenient reference during subsequent data analyses.

In order to use the new ECB technology, the following are required:

- a NELS:88 Compact Disc;
- a CD-ROM reader, used to read or copy the NELS:88 CD-ROM to a personal computer;
- an IBM-compatible personal computer (PC), minimally a 286 system;
- up to 10 Mb space on the PC for the full ECB system;

and,²¹

- a substantial amount of space for the data files. Although up to 165 Mb is required for all publicly-available base year, first follow-up and second follow-up data sets, it is not necessary to copy and/or analyze all of these files simultaneously.

The NELS:88 Compact Disc includes installation procedures, programs and files required by the codebook system, the raw data files and data user manuals (in WordPerfect format).

Different Versions of the ECBs. Table 7.3.2-1 lists three versions of the NELS:88 ECBs that have been created for NELS:88.

The base year school sample is representative of all schools in the nation enrolling eighth graders in 1988. On the first follow-up ECB which includes base year files, information reflecting these schools has been released at two levels of analysis: aggregated at the level of the school (one record for each school), as well as distributed at the level of the students who attended those schools (one record for each such student). However, the second follow-up ECB only includes the base year school data at the level of the student.

The 1994 release of the first follow-up data contains minor adjustments to the cases that are included on the files. For example, sample members found to have been sampled into the study in error have been deleted, and base year ineligible students found to be eligible in the first or second follow-up have been added. A few of the first follow-up variables have also been updated for the second follow-up release of the first follow-up data. Such adjustments are possible in longitudinal studies as new information becomes available or technical advancements become feasible.

Although Table 7.3.2-1 includes both the interim and final versions of the second follow-up CD-ROM, this manual primarily discusses the contents of the final version of the second follow-up CD-ROM. The final second follow-up ECB encompasses thirteen of the major component files through the second follow-up of NELS:88. (The fourteenth major component dataset, the transcript files, appears on the final restricted-use CD-ROM that is not in the ECB format.) Cognitive test variables on all three waves of the survey have been refined and the first follow-up cases have been enhanced by the deletion of ineligibles and the addition of survey-eligible

Table 7.3.2-1

²¹ Space requirements will vary by the ECB component that is selected, the number of variables that may be chosen for generation of a hardcopy codebook, and by the statistical package used by the researcher.

Three versions of the NELS:88 electronic codebooks

<u>ECB Version</u>	<u>Survey Waves and Components</u>	<u>User Version</u>
First Follow-Up ECB	base year, first follow-up (all components)	public use only
Second Follow-Up Interim ECB	base year, first follow-up, and second follow-up, (student and dropout)	public use only
Second Follow-Up Final ECB	base year, updated first follow-up, and second follow-up (student, dropout, school, parent, and teacher) ²²	public use and restricted use

BYI sample members. Both the restricted use and public use CD-ROMs display a weighted and an unweighted frequency window.

A number of restricted-use ASCII files are also available on a separate CD-ROM; these files are not in electronic codebook format. These files include 1) the transcript component data file, data file user's manual, and files of SAS and SPSS control cards for transcript data, 2) all first follow-up and second follow-up School Effectiveness Study data files and control cards, 3) the second follow-up early graduate student supplement, 4) the cognitive test item file which is also on the restricted use CD-ROM that contains the ECB, 5) the expanded sample file, and 6) selected zip code-level community contextual variables drawn from the 1990 Census files for NELS:88 schools. Contents of this CD-ROM are more fully described in the *NELS:88 Second Follow-Up Final Technical Report*.

Magnetic tape versions of the public use data can be ordered from the U.S. Department of Education, Information Technology Branch at (202) 219-1522. The NELS:88 public use data on ECB/CD-ROM, which includes documentation for the ECB, can be ordered by calling Peggy Quinn at (202) 219-1743. The ECB is a qualitative advance over older approaches to complex data sets. The ease with which the pre-analysis phase is handled by the ECB is expected to increase both the number and types of users drawn to the NELS:88

²² The second follow-up restricted use CD-ROM contains an ASCII file of the student component cognitive test items; however, these items are not in the ECB format.

database and, consequently, the variety of research topics addressed. Additional development of the ECB concept is expected to add useful enhancements. Critiques and suggestions on the ECB, the hardcopy codebook and other elements of the NELS:88 data user manuals are welcome. Please address your comments to:

Peggy Quinn
U.S. Department of Education
Office of Educational Research and Improvement
National Center for Education Statistics
555 New Jersey Avenue, N.W.
Room 410H
Washington D.C. 20208
Fax (202) 219-1728

NELS:88 restricted use data on magnetic tapes and on CD-ROM are available at no charge on a restricted loan basis to individuals and/or institutions that obtain an approved license agreement from NCES. To request a license agreement, the individual and/or institution must provide the following information:

- The title of the survey to which access is desired.
- A detailed discussion of the statistical research project that necessitates accessing the restricted NCES survey data.
- The name and title of the most senior official having the authority to bind the organization to the provisions of the license agreement.
- The name and title of the principal project officer who will oversee the daily operations.
- The number, name, and title of professional and technical staff who will access the survey data base. Each professional or technical staff member with access to the data is required to sign and have notarized an affidavit of nondisclosure.
- The estimated loan period necessary for accessing the NCES survey data base.
- The desired computer product specifications, such as medium (9-track tape, CD-ROM, PC diskette), code convention (ASCII, EBCDIC, SAS), etc.

To obtain further details and a license agreement form please write to:

Alan W. Moorehead
Data Security Officer
Statistical Standards and Methodology Division
U.S. Department of Education
Office of Educational Research and Improvement
National Center for Education Statistics
555 New Jersey Avenue, N.W.
Room 408
Washington D.C. 20208
ph. (202) 219-1920