DOCUMENT RESUME

ED 141 406 95 TM 006 369

AUTHOR	Tabler, Kenneth
TITLE	National Longitudinal Study of the High School Class of 1972: First Followup Survey Design, Instrument Preparation, Data Collection and File Development.
INSTITUTION	National Center for Education Statistics (DHEW), Washington, D.C.
SPONS AGENCY	Office of the Assistant Secretary for Education (DHEW), Washington, D.C.
REPORT NO	NCES-77-262
PUB DATE	77
CONTRACT	OEC-0-73-6666
NOTE	48p.; For related documents, see ED 110 730, 117 155, 120 202, 120 251 and 133 334 ; Some parts may be marginally legible due to small print of the original document
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Data Analysis; *Data Bases; Data Collection; Followup Studies; *High School Graduates; High Schools; Information Utilization; Interviews; *Longitudinal Studies; Models; *National Surveys; Post Secondary Education; Questionnaires; *Research Design; Research Methodology; Sampling; Seniors; Test Construction; Weighted Scores
IDENTIFIERS	*National Longitudinal Study High School Class 1972

ABSTRACT

Activities involved in the collection and assembling of data for computer processing from the first followup survey of the National Longitudinal Study of the High School Class of 1972 (NLS) are briefly described. Included are an overview of the NLS; the sample design and survey participation; the development of the first followup survey questionnaires; the means used for maximizing participation and response; the preparation of the NLS Data File; and the calculations of score weights. (MV)

SPONSORED REPORTS SERIES

FIRST FOLLOWUP SURVEY DESIGN, INSTRUMENT PREPARATION, DATA COLLECTION, AND FILE DEVELOPMENT

NATIONAL LONGITUDINAL STUDY OF THE HIGH SCHOOL CLASS OF 1972

NCES 77-262

NATIONAL LONGITUDINAL STUDY OF THE HIGH SCHOOL CLASS OF 1972

FIRST FOLLOWUP SURVEY DESIGN, INSTRUMENT PREPARATION, DATA COLLECTION, AND FILE DEVELOPMENT

Project Officer

Kenneth Tabler

National Center for Education Statistics

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Joseph A. Califano, Jr., *Secretary*

Education Division
Mary F. Berry, *Assistant Secretary for Education*

National Center for Education Statistics
Marie D. Eldridge, *Administrator*

NATIONAL CENTER FOR EDUCATION STATISTICS

"The purpose of the Center shall be to collect and disseminate statistics and other data related to education in the United States and in other nations. The Center shall . . . collect, collate, and, from time to time, report full and complete statistics on the conditions of education in the United States; conduct and publish reports on specialized analyses of the meaning and significance of such statistics; . . . and review and report on education activities in foreign countries."--Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e-1).

Prepared under contract No. OEC-0-73-6666 with the U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects are encouraged to express freely their professional judgment. This report, therefore, does not necessarily represent positions or policies of the Education Division, and no official endorsement should be inferred.

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1977

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Stock No. 017-080-01723-7

FOREWORD

This report is a brief description of the activities involved in the collection and assembling of data for computer processing from the first followup survey of the National Longitudinal Study of the High School Class of 1972 (NLS). The survey was initiated in October 1973 with the mailing of a questionnaire followed by various reminder communications and tracing procedures. At the end of approximately 3 months, questionnaire data on the mail nonrespondents were, to the extent possible, obtained by personal interviews. The success of these efforts is attested to by the response rate of 93 percent.

The initial byproduct of the involvement of the NLS respondents is the conversion of their questionnaire responses and the development of a longitudinal data base for use by the many policymakers and researchers who will be needing and analyzing these and subsequent data collections for years to come. The process by which the study accomplished its objectives and succeeded in compiling the resultant data file is reported herein.

Longitudinal Studies Branch

Elmer F. Collins, *Acting Director*
Division of Multilevel Education Statistics

CONTENTS

FIGURES

TABLES

Chapter I

OVERVIEW: THE NATIONAL LONGITUDINAL STUDY OF THE CLASS OF 1972

A. Data Needs of Policymakers and Researchers

In 1968 the National Center for Education Statistics (NCES) conducted a survey to determine the data needs of educational policymakers and researchers. Survey responses indicated needs for data that would allow student educational-vocational experiences to be compared with later educational-occupational outcomes. This finding provided the impetus for NCES to begin planning the first of a series of national longitudinal studies.

Although the educational-vocational development of young people after high school has been studied, necessary information is lacking for at least three reasons. First, social and economic factors change over time and affect drastically the educational-vocational progress of young people; for example, the Project Talent[1] longitudinal survey of the early 1960's is not relatable to the open-door colleges, modal proportions of minorities entering colleges, etc., of the 1970's. Second, many studies are concerned with only one class of antecedent variables for explaining later developments, and thus fail to take account of interventions emanating from current Federal priorities and to include representative samples with sufficient numbers of important subgroups (e.g., racial-ethnic minorities). Third, many studies are primarily concerned with developing models or theories of behavior (e.g., Donald Super's Career Pattern Study[2]) an objective not alien to but insufficient for the objectives of the National Longitudinal Study of the Class of 1972 (NLS). Behavioral studies have given, however, an excellent basis for perfecting the NLS design. For indeed, the continuing planning has relied considerably on the review and synthesis of the findings of studies that apply to educational-vocational development, such as the one by UCLA's Evaluation Center[3] for the U.S. Office of Education (OE).

In April, 1970, educational researchers and administrators met with Federal officials in Washington, D.C. The NLS reflects their guidance and the data needs of NCES and several OE agencies: the Office of Planning, Budgeting, and Evaluation; the Bureau of Postsecondary Education; the Bureau of Occupational and Adult Education; and the Bureau of Education for the Handicapped. Four advisory committees guided the NLS planning. One committee was composed of research experts and representatives of educational organizations; two others were made up of officials of State education agencies; and the fourth, an internal OE user's committee, represented offices and bureaus of the Department of Health, Education, and Welfare (HEW).

Later in 1970, the basic survey planning was contracted to Research Triangle Institute (RTI) and the sample planning and design to WESTAT, Inc.[4] After extensive planning, which included the design and field testing of survey instrumentation and procedures, the first full-scale NLS survey was initiated in the spring of 1972.

B. Data Collections, 1972-73

In 1972, a national probability sample of more than 18,000 seniors from 1,070 public, private, and church-affiliated high schools participated in the base-year

[1] John C. Flanagan and William W. Cooley. 1966. *Project Talent One-Year Follow-Up Studies.* University of Pittsburg.

[2] E. D. Super, R. S. Kowaski, and E. M. Gotkin. 1967. *Floundering and Trial After High School.* Career Pattern Study Monograph IV. New York: Columbia University.

[3] James W. Trent, Claire Rose, Ann Salyard, Judd Adams, Alfred C. Marcus, J. Ward Keesling, and Arthur Gerst. 1972. *An Analytical Review of Longitudinal and Related Studies as They Apply to the Educational Process.* Los Angeles: Center for the Study of Evaluation, UCLA.

[4] WESTAT, Inc. 1972. *Sample Design for the Selection of a Sample of Schools With Twelfth-Graders for a Longitudinal Study.* Rockville, Maryland.

survey conducted by Educational Testing Service (ETS).[5]

1. Base-Year Student Survey Instruments

Each student in the sample was asked to complete a Student Questionnaire containing 104 questions distributed over 11 sections. The questions dealt with personal-family background, educational and work experiences, plans, aspirations, attitudes, and opinions. Students were given the option of completing the questionnaire in school or taking it home to get assistance from their parents. Those selected for the survey were informed in the questionnaire and in a newsletter of the voluntary nature of participation, of their prerogative to skip questions they considered personally sensitive; and of the objectives and importance of the study for future educational benefits. Participants were assured that their responses would be treated as confidential, that they would remain anonymous, and that data collected by the survey would be published only in aggregate form.

Each student was also asked to complete a 69-minute Test Book designed to measure both verbal and nonverbal abilities. The book contained six tests which are described briefly below in the order of administration:

Vocabulary. A brief test using synonym format. The items were selected to avoid academic or collegiate bias and to be on an appropriate level of difficulty for the 12th-grade population. (15 items, 5 minutes)

Picture number. A test of associative memory consisting of drawings of familiar objects, each paired with a number. The student, after studying the picture-number pairs, was asked to recall the number associated with each object. (30 items, 10 minutes)

Reading. A test based on 100-200 word passages with questions concerning various reading skills (analysis, interpretation) and focused on straightforward comprehension. With the vocabulary test, it provided a means to derive a verbal score which could allow links to the normative data available for the SAT. (20 items, 15 minutes)

Letter groups. A test of inductive reasoning requiring the student to draw general concepts from sets of data or to form and try out hypotheses in a nonverbal context. The items had five groups of letters; four groups shared a common characteristic. The student indicated which group differed from the others. (25 items, 15 minutes)

Mathematics. Quantitative comparisons requiring the student either to indicate which of two quantities is greater or to assert equality or the lack of data for comparing. This item was relatively quickly answered and provided measures of basic competence in mathematics. (25 items, 15 minutes)

Mosaic comparisons. A test of perceptual speed and accuracy with items requiring that small differences be detected between pairs of otherwise identical mosaics or tile-like patterns. A deliberately speeded test, it had three separately timed sections of increasingly complex patterns. (116 items, 9 minutes)

From each student's School Record Information Form (SRIF), data were obtained on the high school curriculum, grade point average, credit hours in major courses, position in ability groupings (if applicable), remedial-instruction record, involvement in certain federally supported programs, and scores on standardized tests.

2. School and Counselor Survey Instruments

Two other data collection instruments were the School Questionnaire and the Counselor Questionnaire. Survey administrators completed the School Questionnaire, which provided information on –

Programs and students. Grade structure, enrollment by curriculum, programs for the handicapped and disadvantaged, teaching, absence and dropout rates, racial-ethnic makeup, college recruitment efforts.

Resources. Participation in Federal programs, teacher turnover, percentage of teachers with advanced degrees, library and other facilities, age of buildings, nearness to postsecondary institutions

Grading systems. Form of the system in use, plus a table of grade equivalents

A maximum of two counselors in each school filled in

[5] Educational Testing Service. 1973. *The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972–Final Report*. Princeton, New Jersey.

the Counselor Questionnaire with data on training, experience, activities, assignments, methods, workload, and resources.

3. Followup Surveys

In the summer of 1973, 4,439 students who did not participate in 1972 were contacted ("resurveyed") to be added to the planned followup sample.

In October of 1973, the first followup survey was begun by RTI. Of the 22,654 young adults expected to participate, 94.2 percent completed the first followup instruments: 61.8 percent by mail and 32.4 percent by personal interview. Participants were asked in First Followup Questionnaires where they were in October 1973 and what they were doing with regard to work, education, and/or training. Similar questions were also asked for the same period in 1972 to facilitate the tracing of progress since high school and the defining of factors that affect that progress. The first followup data collection was completed in April of 1974. (Content of the First Followup Questionnaire is covered in detail in chapter III of this report; first followup data collection activities are discussed in chapter IV.)

The second followup survey was begun by RTI in October 1974, with data collection completed in April of 1975. A third followup of the class of 1972 was begun in the fall of 1976.

C. Uses for NLS Data

Periodically, data are being obtained from members of the class of 1972 and added to their individual histories that is, to their experiences, activities, attitudes, satisfactions, environments, and plans as they move into the critical years of early adulthood. These data will fill widespread needs of the educational community researchers and administrators in the elementary, secondary, and postsecondary educational-occupational systems. The data will provide insights into identifying and understanding the major branching or decision points that affect the educational and life patterns in the immediate postsecondary period. Significant linkages of path choices can be traced, associated transition probabilities can be estimated, and insight into the relative importance of factors which determine these probabilities can be realized.

1. To Clarify Choices and Alternatives

Collectively, the individual histories should provide quantitative data for policymakers, planners, and researchers about various issues

- The demands for postsecondary education and training, including vocational/technical education
- The abilities and characteristics of actual and potential users of postsecondary education
- The extent to which earlier plans and aspirations persist over time and are eventually fulfilled
- The reasons why young adults change their plans and fail to accomplish earlier objectives
- The impact of federally funded postsecondary programs on initial choices and later activities and plans
- The factors influencing young people in choosing their lifework and in determining success and satisfaction in this work
- The extent to which educational experiences have prepared young people for their work
- The characteristics and abilities of those making occupational choices and the reasons why choices are made
- The impact of high school experiences, curriculums, peer-group aspirations, guidance counseling, etc., on initial educational and occupational plans and on perseverance and success in achieving them
- Young adults' awareness of educational and occupational alternatives, their perceptions of options open to them, and the extent to which they have been limited by lack of information
- Financial characteristics in setting low-aspiration goals and in failing to meet high-aspiration goals

2. To Trace Progress

The primary NLS purpose is to discover what happens to young people after they leave high school and to relate this information to their prior educational experiences and their personal and biographical characteristics. Ultimately, the study will lead to a better

understanding of the development of students as they pass through the American educational system and of the complex factors associated with individual educational and career outcomes. Such information is essential as a basis for effective planning, implementation, and evaluation of Federal policies and programs designed to enhance educational opportunities and achievements and to upgrade occupational attainments and career outcomes.

3. To Provide a Data Base

The major NLS objectives are to provide a data base for policy decisions that may guide Federal contributions through the Nation's educational system to the fullest development of human and material resources, and to provide social scientists and scholars with a rapidly enriching data base that no one of them could afford but all can use in pursuit of professional interests. The data base itself, as it now exists and as it is updated, will be available to interested researchers for their own studies.

4. To Disseminate Information

Summaries and analyses of data are being written up in periodic reports to be issued by HEW and made available to the educational community. Reports focused on single issues targeted for specific groups will be designed for rapid dissemination.

D. Objectives of the First Followup

The 1973-74 first followup survey included data collection, data transformation, preliminary data analysis and interpretation, and the making of survey recommendations for the 1974-75 second followup. The overall aim of these activities was to satisfy the broad, long-range NLS objectives:

1. To assess the demands for postsecondary education, including adult, vocational, and technical; the characteristics of students going on; where they go; the factors inhibiting the realization of educational aspirations; and the characteristics of and alternatives pursued by those who do not go on.

2. To determine what types of students make what educational and/or occupational choices--for the purpose of establishing meaningful flow data, understanding the chain of decisions that shape an individual's education, training, and launching of a career, and establishing the relationships needed for predictions.

3. To develop means for assessing how educational experiences, personal influences, and social attitudes have led the graduate to the point at which he finds himself, and for evaluating the extent to which these are related to his decisions about occupational choice, military service, and vocational, technical, and higher education.

4. To determine the extent to which students are aware of postsecondary education and employment alternatives and the options still open to them; how these were influenced by counseling; to what extent institutional and federally funded recruiting projects affect postsecondary school choices; how employment opportunities can influence these choices; and to what extent the lack of information about postsecondary opportunities (either education or occupational) limits their aspirations.

5. To determine the influence of student ability on postsecondary choices and to associate the choices with test scores, the student's perception of his own abilities, and his class standing.

6. To relate low-aspiration choices for postsecondary education to the principal obstacles perceived by the student, especially financial obstacles, and to determine profiles of the student's knowledge of programs of financial aid, their applicability to his situation, and his intention to take advantage of them.

7. To follow the educational progress of students and those terminating early to see how high school experiences, curriculum patterns, and financial and other factors are associated with postsecondary career choices and perseverance and success in them, and to identify the factors associated with "dropping out" and changing jobs after different intervals.

8. To identify, from the cohort study, subpopulations such as high achievers with limited financial resources, disadvantaged minority groups, and students in junior colleges and vocational and technical schools, and to investigate interactions and influences between and among individuals that will shape their futures.

9. To refine the means and methods of assembling, merging, and maintaining data on large, diverse samples of highly mobile populations and to relate these techniques to other fields.

10. To investigate the consistency of patterns apparent in the analysis of base-year data, identify new patterns and/or changes in established patterns for further investigative emphasis, and define areas for emphasis in subsequent follow-ups.

Chapter II

SAMPLE DESIGN AND SURVEY PARTICIPATION

A. Basic Design

All eligible 1972 12th-graders enrolled in all public, private, and church-affiliated high schools in the 50 States and the District of Columbia made up the population to be sampled for the NLS base year. The survey plan for selecting from geographical areas was a deeply stratified two-stage probability sample design with schools as first-stage sampling units and students as second-stage units. The sample was designed and selected by WESTAT, Inc.[1]

The school sampling frame, constructed from computerized school files maintained by OE and by the National Catholic Education Association, was stratified into 600 final strata based on the following variables:

- Type of control (public or nonpublic);
- Geographic region (Northeast, North Central, South, and West);
- Grade 12 enrollment (< 300; 300-599; > 600);
- Proximity to institutions of higher learning (3 levels);
- Percentage minority group enrollment (8 levels);
- Income level and type of community (2 types); and
- Degree of urbanization (10 levels).

Schools in low-income communities and schools with high proportions of minority group enrollments were selected and assigned probabilities twice as large as those used for the other schools to increase the numbers of disadvantaged students in the sample. Schools in the smallest grade enrollment stratum (<300 seniors) were selected with probabilities proportional to their estimated numbers of seniors and without replacement; schools in the remaining grade 12 enrollment strata were selected with equal probabilities and without replacement. Within each of the 600 strata, 4 schools were selected; then 2 of the 4 were designated as the primary selections (2 x 600 = 1,200) and the other 2 were retained as backups (1,200) to be used in the sample if one or both of the primary schools did not cooperate (e.g., refused, were ineligible). From each school, 18 students and 5 alternate students were sampled with equal probabilities and without replacement.

B. Base-Year Participation, 1972

The sample design involved 1,200 primary sample schools and a target sample of 21,600 students (18 per school). The task of collecting base-year data was contracted by NCES to Educational Testing Service (ETS) of Princeton, New Jersey.[2] Of the 1,200 schools, 948 participated, 21 had no senior students enrolled, and 231 either refused to or could not participate because the request was received too late in the school year (table 2-1). A participating school was defined as one in which at least one Student Questionnaire, Test Book, or Student Record Information Form (SRIF) was completed.

Table 2-1.–Base-year school participation data

Sample schools	Primary sample schools		Participating schools	
	Number	Percent	Number	Percent
Total	1,200	100.0	1,070	100.0
Primary sample				
Participants	948	79.0	948	88.7
No senior students	21	1.8	---	---
Nonparticipants	231	19.2	---	---
Backup sample				
"Extra" in base year	---	---	26	2.4
Other participants	---	---	96	8.9

[1] WESTAT, Inc. 1972. *Sample Design for the Selection of a Sample of Schools With Twelfth-Graders for a Longitudinal Study*. Rockville, Maryland.

[2] Educational Testing Service. 1973. *The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972–Final Report*. Princeton, New Jersey.

Backup schools were also established as a part of the sample design as replacements for nonparticipating primary sample schools. In the base year, 122 backup schools participated; 26 of these were termed "extra" because both primary sample schools from that stratum also participated. In the 1972 survey, there were 1,070 participating schools.

Within the 1,070 schools, 19,144 students were sampled; these included the 18 eligible students per school and the alternates used to replace noneligibles. Of the 19,144, 95 percent participated (table 2-2) by having available one or more completed survey instruments. The average was 17 participants and 1 nonparticipant per school.

Table 2-2.--Base-year student participation data

Sample students	Number of students in 1,070 schools		
	Total	Average per school	Percent
Total	19,144	17.91	100.0
Participants	18,143	16.97	94.8
Nonparticipants	1,001	0.94	5.2

Response rates (table 2-3) varied by instrument--SRIF's were completed by 95 percent of the 19,144 sample students, Student Questionnaires by 87 percent, and Test Books by 83 percent. Looking at the response rates another way, SRIF's were filled out for all but 33 of the 18,143 participating students--nearly a 100 percent response--and Student Questionnaires and Test Books were completed by 92 percent and 87 percent.

Table 2-3.--Base-year student response rates

Survey instrument	Number completed	Percent of sample students*	Percent of participating students**
SRIF	18,110	94.6	99.8
Student Questionnaire	16,683	87.1	92.0
Test Book	15,863	82.9	87.4

*19,144 in 1,070 participating schools.
**18,143 in 1,070 participating schools.

The School Questionnaire was completed by 97 percent (1,038) of the participating schools; the Counselor Questionnaire was completed by 1,804 counselors in 1,040 schools.

C. Resurvey Plans

Due to the large school nonresponse in 1972, it was necessary to elicit responses from the 231 nonparticipating primary schools and replacements for the 21 schools that had no seniors. The initial "resurvey" activity conducted in the summer of 1973 before the first followup survey of October 1973 involved securing school cooperation and selecting random samples of up to 18 students (1972 seniors) per school. RTI's basic responsibilities were to:

1. Collect School Questionnaire data from each resurvey school;
2. Collect SRIF data on individuals selected from each resurvey school, secure permission to obtain the same SRIF data from individuals selected for resurvey;
3. Collect short Base-Year Questionnaire data from those individuals selected from each resurvey school;
4. Update the base-year data base and produce weights to incorporate the additional data.

The information gathering was limited to relatively hard data rather than data on attitudes, plans, and perceptions likely to change substantially over time.

D. Resurvey Participation, 1973-74

In the summer of 1973 when NCES contacted the 231 sample schools that had not participated in the base-year survey, most of them (205) provided the latest names and addresses of up to 18 of their 1972 seniors selected by simple random sampling. Former 1972 senior students were also selected from 16 augmentation sample schools (in 200 school districts, eligible public schools not included in the base-year sampling frame). Thus 4,315 former students were added from 257 participating resurvey schools--205 primary sample, 36 backup sample, and 16 augmentation sample schools.

A newsletter developed by RTI was mailed in July 1973 not only to encourage participation but also to update names and addresses. When mail was returned by the U.S. postal service as undeliverable, telephone tracing procedures were used to obtain current addresses. Table 2-4 shows the response rates of the resurvey group. The individual response rate to Form B was acceptably high; however, school cooperation in

completing School Questionnaires and providing SRIF's proved difficult to obtain. SRIF's were requested from schools only for those resurvey individuals who gave permission for the data to be used.

Table 2-4.--Resurvey student response rates

Survey instrument	Number completed	Number in sample	Percent participating
Form B. Questionnaire*	4,061	4,315	94.1
School Questionnaire	202	257	78.6
SRIF-Use Permission	3,467**	4,315	80.3

*With base-year background data.
**Includes 880 SRIF's received after cutoff date of first release tape.

E. First Followup Participation, 1973–74

Table 2-5 shows the numbers of schools participating in the base year, the resurvey, and the first followup. Not included in the first followup were 26 schools termed "extra" in the base year. Due to the resurvey, 1,153 or [redacted] percent of the 1,200 primary sample schools did participate. With the backup and the augmentation sample, the first followup participation was 1,300 schools.

Table 2-5.--Number of participating schools, by survey

School sample	The base-year survey	The resurvey effort	First followup survey
Total	1,070	257	1,300*
Primary sample	948	205	1,153
Backup sample			
"Extra" in base year	26*	...	...
Other	95	36	131
Augmentation sample	...	16	16

*The 26 extra schools were not included in the first followup.

Table 2-6 shows the composition of the first followup sample membership. Included in the final NLS sample of 23,451 are 19,136 base-year participants and 4,315 resurvey members. Base-year extras not surveyed in the first followup, removals from the file, and participants with bad addresses, etc., yielded a final followup mailout sample of 22,654.

Table 2-6.--Composition of first followup sample membership

A. *Received from prior contractor*			
Base-year respondents from:			
Primary sample schools		18,502	
With Student Questionnaires	16,301		
Without Student Questionnaires	2,201		
Backup sample schools		499	
With Student Questionnaires	382		
Without Student Questionnaires	117		
Other schools		135	
Without Student Questionnaires	135		
Total sample received		19,136	19,136
Duplicate and ineligible records (later removed)		8	
Total records received		19,144	
B. *Resurvey sample*			
Original sample school members		4,047	
Augmentation sample school members		268	
Total resurvey sample			4,315
C. *Total NLS sample*			23,451
D. *Presurvey removals*			
Base-year extras		− 472	
Refusals		− 34	
Deceased		− 5	
Transfers or dropouts prior to graduation		− 21	
Miscellaneous (EMR, out of country, etc.)		− 9	
Incomplete or garbled addresses, untraceable, etc.		− 256	
Total removals			− 797
E. *Total first first followup mailout*			22,654

Chapter III

DEVELOPMENT OF THE FIRST FOLLOWUP QUESTIONNAIRES

A. Modifying a Model

One early effort in planning for the NLS was a large-scale literature review and analysis.[1] Selected major longitudinal studies of student and nonstudent populations were analyzed to discover variables, techniques, methodologies, and problems pertinent to evaluations of the effects of schools and colleges on children and young adults between the ages of 14 and 30.

One product of this effort was the conceptualized General Educational Development Model depicted in figure 3-1. This model was used in the planning for the First Followup Questionnaire; it was later modified to include the Base-Year Student Questionnaire.

The modification excludes some classes of variables (e.g., psychological climate and socialization in the home) since they were not considered amenable to adequate measurement. Other classes (e.g., personal, intellectual, and educational satisfactions) explicitly included in the original model were later considered to be best measured in future followups. Others were combined; for example, goal orientation in the original model encompasses expectations and aspirations in the later model. Still others (e.g., community environment) not explicitly included in the first model became key to the later model. The modified model is shown as figure 3-2.

B. Defining the Questionnaire Objectives

The questionnaire was designed to meet the primary analytical objectives of the first followup survey by providing continuity over time in measurements of educational-vocational plans of the base-year respondents. A programmatic objective was to identify sociocultural variables which may have had impacts on 1972–73 educational-occupational decisions. In view of these two broad objectives, questionnaire items were designed to describe the flow pattern between an individual's activity statuses in October 1972 and October 1973 and the future plans, aspirations, opinions, and attitudes relevant to educational-vocational outcomes. To guide item selection and revision, 10 questionnaire objectives were defined.

OBJECTIVE 1—To document the activity statuses and programs in October 1972 and October 1973.

Student activities and progress in 1972–73 were determined by requesting information beyond that asked for in item 1: "What are you doing *now*?" Groups of items follow:

EDUCATIONAL STATUS

School attendance
Educational and training programs since high school
Postsecondary school attended since high school
Classes or courses in October 1973
Classes or courses in October 1972
Other school attended since high school
Current attendance at other school

Institution type
Name, location, and type of school in October 1973
Name, location, and type of school in October 1972
Name, location, and type of other school

Program type
Type of educational or training program
Field of study or training in October 1973
Field of study or training in October 1972

Program completion
Completion of educational or training program
Grades

[1] James W. Trent, Claire Rose, Ann Salyard, Judd Adams, Alfred C. Marcus, J. Ward Keesling, and Arthur Gehrt. 1972. *An Analytical Review of Longitudinal and Related Studies as They Apply to the Educational Process.* Los Angeles: Center for the Study of Evaluation, UCLA.

Figure 3-1.--General educational development model

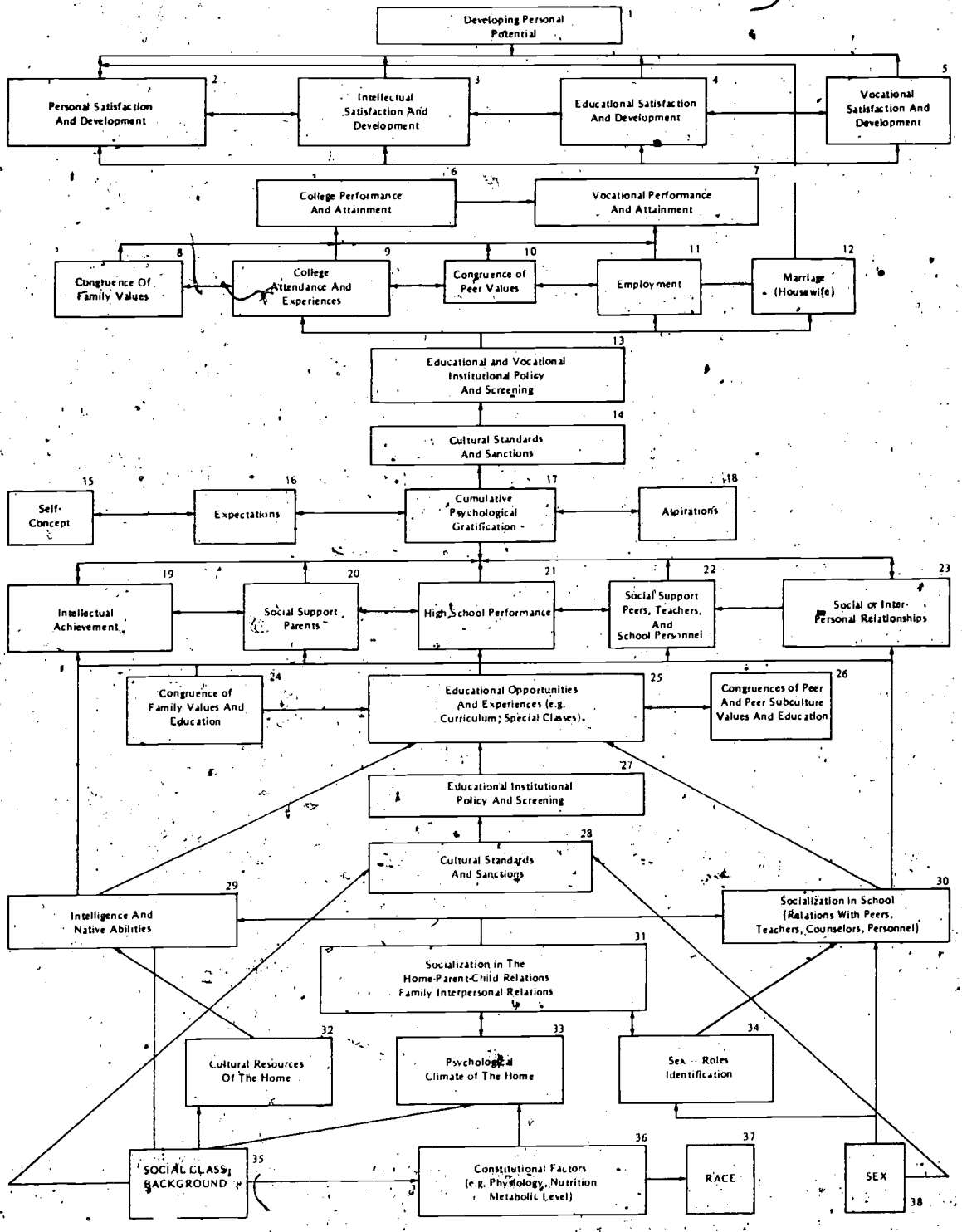
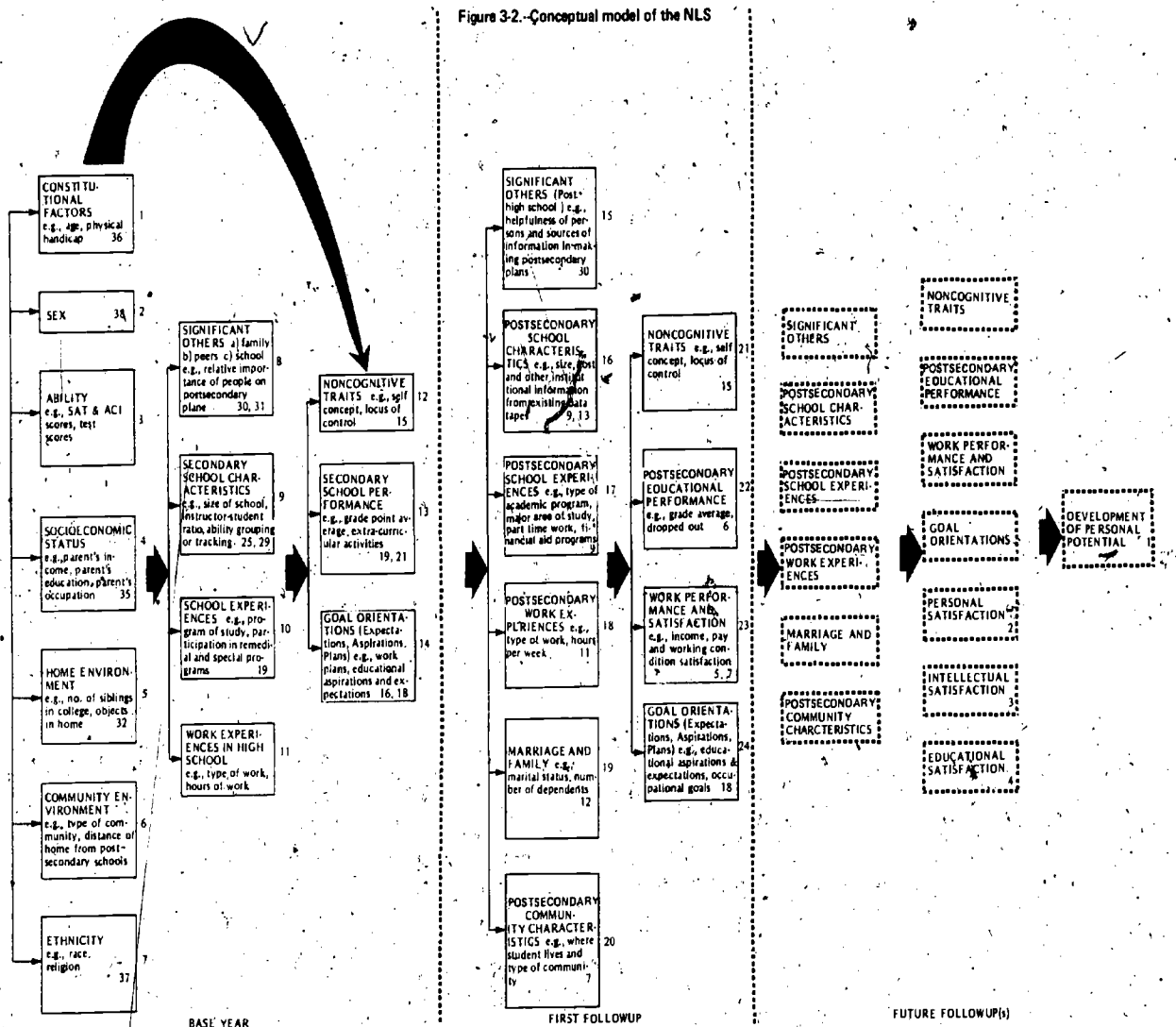


Figure 3-2.--Conceptual model of the NLS



Program completion–Continued.
Credits earned
Certificate, degree, or license working toward
Certificate, degree, or license earned

EMPLOYMENT STATUS

1973 and 1972
Working in October
Reasons not working in October
Looking for work in September
Job description in October
Hours worked per week in October
Earnings per week in October
Other jobs in October

MILITARY STATUS

Current status
Branch of Armed Forces
Enlisted or drafted
Date began active duty
Specialized schooling
Type of program
Pay grade and specialty rating
Courses taken
Active duty or date left

MARITAL STATUS

Marital status
Marriage date

OBJECTIVE 2–To identify background factors associated with postsecondary school choices and career paths and persistence in those choices and paths.

Background factors (input variables) considered to be important predictors of postsecondary school choices and career paths were measured in the Base-Year Questionnaires. Most of them did not need to be repeated; however, the first followup survey was to include about 4,300 individuals not surveyed in the base year (the resurvey members, for whom an augmented form of the questionnaire was prepared). Additional base-year questions on this augmented form included socioeconomic status and home environment; race or ethnicity; religion; community characteristics; influences of significant others such as peers, parents, and school personnel; high school program and activities; and personal characteristics such as ability, sex, age, and physical handicaps.

OBJECTIVE 3–To identify contemporaneous variables associated with persistence and changes in plans in spring of 1972 and activity statuses in October 1972 and October 1973.

The purpose was to estimate the effects of contemporaneous events–those intervening events occurring after a student leaves high school–upon persistence in activity statuses between October 1972 and October 1973. (As a rule, it is necessary to control key background variables such as socioeconomic status in order to assess the effects of postsecondary school variables.) Variables which research has suggested may be important for facilitating or inhibiting a student's progress after leaving high school are:

Economic factors are measured by items 9 and 10 on the contemporaneous financial independence and responsibilities of the individual's parents and himself, item 11 on individual's and spouses' incomes, items 46 and 47 on costs of and sources for paying for education.

School experiences are measured by items 24 and 29b on reasons for not continuing education; items 81-84 on the college applications; items 27a, 33a, and 40c on matriculation dates; items 27b and 33b on full- versus part-time status; item 31 on reasons for changing schools; item 38 on reasons for withdrawing from school; items 22, 28, and 36 on fields of study; items 34 and 35 on changing fields; item 44 on participation in special services programs.

Work experiences are measured by items 48b and 54b on reasons not working, item 51 on job satisfaction, item 59 on methods used in looking for work, items 62 and 63 on the relatedness of high school vocational training and subsequent work experience.

Marriage and family planning are measured by item 7 on marital status; items 8, 17, and 18 on fertility behavior and plans; items 24, 29b, 48b, and 54b on marriage (homemaking) as a reason for not working or going to school.

Community characteristics are measured by item 6b

on migrations, items 4 and 5 on people lived with and type of residence, item 6a on location (suburb, farm, city, etc.).

Psychological traits are measured by item 15 on self-concept and control of environment, item 13 on willingness to borrow money for education.

Influences are measured by items 77 and 85 on encouragement, discouragement, and helpfulness of significant others in making vocational-educational decisions.

OBJECTIVE 4–To measure shifts in general and specific life goals in short- and long-range planning.

Plans or aspirations are not always congruent with the states in which individuals find themselves at any given point in time. Thus, in October 1973, we wanted to know their current plans and future goals. Changes in goals since they were seniors were examined to see if the changes were in a direction consistent with educational-vocational qualifications and/or experiences. To maintain comparability, the majority of items measuring aspirations in the First Followup Questionnaire were repeats of items in the Base-Year (B-Y) Student Questionnaire.

Item 20 measures life goals (as did B-Y item 20); items 12 and 14 measure educational aspirations and expectations (same as B-Y item 29); item 19 identifies the kind of work they expect to be doing when they are 30 (revision of B-Y item 25); items 75 and 76 measure plans for military service; and items 7a, 17, and 18 deal with marital and childbearing plans important to predicting educational-vocational outcomes of men and women.

OBJECTIVE 5–To estimate the effects of institutional and federally financed programs on aspirations and persistence in career programs.

Some items help identify the level of special assistance needs of students with handicaps and refer to the uses that students make of different programs.

Items 24 and 29b deal with reasons for not continuing education, items 31 and 35 concern reasons for changing school and field of study, and items 48b and 54 concern reasons for not working and identify students in need of assistance. Individuals who have received special assistance or have used institutional resources may be identified by item 85 on helpfulness of high school counseling programs; items 82d, 83d, and 84d deal with offers of financial aid for continuing education; item 47 concerns sources for paying school and college expenses; item 44 measures participation in special services programs; item 22a deals with on-the-job training and other manpower programs; item 59 concerns the uses of agencies in finding work; and items 72 and 76 deal with uses of the GI bill in making educational plans.

OBJECTIVE 6–To identify subgroups of potential interest for special analysis.

The measures required for special analyses of postsecondary school dropouts (e.g., those of high ability and disadvantaged minority groups) are covered by items in the Student Questionnaire or by items mentioned under the above objectives.

OBJECTIVE 7–To measure short-term progress with respect to the relative effectiveness of high school and postsecondary vocational and on-the-job training.

Whereas many individuals who continued their education after leaving high school will not have completed their program of studies, some students who have received career training will not be in the work force.

The short-term effects of career training are assessed by item 50b on earnings per week and by item 51 on job satisfaction. Items 22 and 43 describe programs completed and certificates earned in postsecondary vocational, business, or trade schools; items 61, 62, and 63 assess the utility of high school vocational training programs.

OBJECTIVE 8–To assess the influence of the high school "environment"–as reflected by the variance in ability and socioeconomic background of the students in a school–on the progress of subgroups in different activity statuses.

This objective may be fulfilled by using the measures of progress outlined under objective 3. Items used as predictors and as the bases for classifying selected subgroups of special interest will be taken from the Base-Year Student and School Questionnaire.

OBJECTIVE 9—To define areas for concentration in subsequent followups in order to assess the effects of education on nonvocational and nonacademic outcomes, including cognitive and noncognitive measures.

While conventional measures such as employment status and income will be used in later analyses to distinguish the "successful" from the "unsuccessful" students, education may have other significant outcomes that often are overlooked. Thus the cognitive traits "self-esteem" and "locus of control" (item 15) and life goals (item 20) were repeated from the Student Questionnaire.

OBJECTIVE 10—To develop a followup questionnaire for assembling, merging, and maintaining data on a large sample of a highly mobile population and to develop the means for clearly defining the target population at any point in time.

The timing of events was critical in the followup since questionnaires would be administered to respondents over a 2- to 6-month period. To provide a common reference date for the entire sample, most recent events are linked to the October 1973 date. However, since psychological traits such as "attitudes" can be measured only contemporaneously (as many as 6 months after October), it will be important to know if any major change has occurred in activity statuses in the interim: Item 1 ("What are you doing *now*?"), items 22c, 28c, and 36c on program length; item 22d on program completion; and items 49g and 55g on current work status detect such changes.

C. Field Testing and Revising the Initial Version

Between the base-year data collection in June 1972 and the start of the first followup in October 1973, the Bureau of the Census field tested the initial version of a full-scale questionnaire approved by the Office of Management and Budget as No. 41-S-73004. This questionnaire was administered during March-May 1973 to 903 high school seniors from the class of 1971—the same sample selected and used in 1971 to design the base-year survey.

The field-tested questionnaire was to be used in the first followup to ascertain the educational and work statuses of the 1972 school seniors for October 1972 and October 1973 and to obtain data on self-concept, locus of control, aspirations, expectations, plans, goals, and work and educational experiences. It was the basis for item revisions and additions which were outgrowths of the application to it of the conceptual models. After 6 weeks of conferences, staff meetings, and draft working papers, the final form of the First Followup Questionnaire was approved in September 1973.

The field-tested instrument contained 63 items. The final form contained 85. The latter instrument was longer because—

1. Time references over the relatively long period (October 1973-March 1974) for collecting data on educational and work activities were addressed to two points in time, October 1972 and the first week of October 1973. (The field test version concentrated primarily on the respondent's current activity.)

2. New areas thought to be of sufficient importance to be included in the new instrument were fertility behavior and plans, life goals, school progress, noncollege education and training programs, and financial independence.

3. Some old questions (e.g., those involving college choice) had to be expanded in scope to obtain the data to answer questions of interest.

4. Father's education and occupation were added to the followup instruments. Many of the proposed first followup analyses were to involve work and educational paths of subpopulations defined in part by family socioeconomic status, and there was substantial nonresponse to these items in the base-year survey.

5. Certain base-year questions needed to be asked of first followup respondents who had not or who had partially participated in 1972 (provided some data but did not complete a Student Questionnaire). Plans were being made to recontact nonparticipant schools and to select about 4,450 students. Both groups would need retrospective base-year questions; Forms A and B of the First Followup Questionnaire were developed.

D. Grouping the Items

Form A of the First Followup Questionnaire is an 85-item, 24-page booklet for self-administration by the respondent. Most of the base-year and first followup

questions are of the fixed-choice (closed-response) types. Open-response questions were limited to dates, income, number of hours or weeks worked, and the like. All questions were constructed and formatted by specialists in instrument design.

Form B of the questionnaire was designed for participants from whom selected base-year data were needed. Identical to Form A through 85 items and 24 pages, Form B continues for 14 items (86-99) in section E, which concerns high school program, activities, and plans; family information; and personal demographic information. Form B also contains a signature block on the inside front cover for permission to use school records. Table 3-1 indicates the first followup and base year equivalents for these 14 items.

Forms A and B are divided into six sections: general information, education and training, civilian work experience, military service, information about the past, and background information.

Table 3-1.–First Followup and Base-Year Questionnaire numbers

First Followup (Form B)	Base-year
86	2
87	5
88	8
89A through 89I	10A through 10I
90	27
91	16
92A and 92B	91A and 91B
93	93
94A through 94K	94A through 94K
95	84
96	88
97	92
98	95
99	83

1. Section A–General Information

Section A contains questions about present activity statuses (1); high school status (2,3); present community and home characteristics (4,5,6); present marital status and fertility (7,8); present financial dependence and income (9,10,11); educational aspirations and expectations, including financial plans (12,13,14); self-concept and locus of control (15); expected future activity statuses, including plans for childbearing (16,17,18,19); and goal orientations (20).

The marital status and fertility questions can be especially useful in explaining the career patterns of men and women. Self-concept, locus of control, life goal orientations, educational expectations and plans were also measured in the base year; thus the absolute values of and the changes in these for the two time points can be hypothesized to moderate the conditional probabilities of leaving one plan or activity state for another.

2. Section B–Education and Training

Section B contains questions about nonschool, nonmilitary training programs (21,22); formal postsecondary education (23); reasons for not continuing formal education (24,29b); 1973 educational status (25,27); characteristics of school attending (26); 1973 field of study or training (28); 1972 educational status (29a,33); transferring and withdrawing (30,31,37,38); characteristics of school attended in 1972 (32); change in field of study or training (34,35); 1972 field of study or training (36); attendance at other schools (39,40); school progress (41,42,43); special services (44); satisfactions with education and training (45); and schooling or training finances (46,47).

The questions in this section were designed primarily for defining postsecondary educational and training activity statuses which are to be used as outcomes in the descriptive, predictive, and causal-modeling analyses.

3. Section C–Civilian Work Experience

Section C contains questions about present employment status (48a,48c); reasons for not working (48b); characteristics of present job (49,50); satisfaction with that job (51); work expectations for 1974 (52); other jobs in 1973 (53); employment status in 1972 (54a,54c); reasons for not working in 1972 (54b); characteristics of 1972 job (55,56); other jobs in 1972 (57); summary of jobs and weeks worked, October 1972 to October 1973 (58); methods of looking for work (59); and employment using high school training (60,61,62,63).

The questions in this section were designed for defining work activity statuses which are to be used as outcomes in descriptive, predictive, and causal-modeling analyses.

4. Section D–Military Service

Section D contains questions about branch and status of military service (64,65,66,67,74), specialized schooling (68,69,71), pay grade and specialty rating

(70), plans to use the GI bill (72), satisfaction with the military service (73), military career plans (75), and postmilitary career plans (76).

The questions in this section are used to define military activity status in descriptive, predictive, and causal-modeling analyses.

5. Section E—Information About the Past

Section E contains questions about influences of parents and friends on postsecondary activities (77); parental educational levels (78); father's occupation (79); mother's employment status (80); prior application, acceptance, and financial aid at first-choice school or college (81,82); second-choice and third-choice schools or colleges (83,84); and helpfulness of high school counseling services (85).

Section E of Form B contains 14 additional items (86–99) which were discussed in this chapter.

6. Section F—Background Information

The inside back pages of Forms A and B contain section F, which asks for the names, addresses, and telephone numbers of the respondent, respondent's parents, and two other persons; and the date of birth, social security number, and driver's license number of the respondent. This information is used in tracing activities for subsequent followup surveys.

E. Selecting the Key-Response Items

Many considerations went into deciding whether a returned questionnaire contained adequate information for acceptance, editing, and entry into the data file. The general requirement was to obtain a certain required minimum of data on respondent activities since leaving high school, rather than complete data on all items. The following items were designated as crucial to the acceptability of the questionnaire:

General Information

1 What are you doing *now*?

16 What do you expect to be doing in *October 1974*?

Education and Training

21 Since high school, any training programs? If "yes," 22a should be completed. (The response to 21 can be presumed if 22a is completed.)

23 Since high school, any colleges or universities? If "yes," 25 should be answered. If 25 is "yes," 26a and 26b should be answered.

29a Any classes or courses in October 1972? If "yes," 30 should be answered. If 30 is "no," 32a should be answered.

37 Withdraw from school prior to completion? If 34 is "no," 37 should be completed.

Civilian Work Experience

48a Holding a job in *October 1973*? If "yes," 49a,b,c,d should be answered.

54a Holding a job in *October 1972*? If "yes," (different job), 5a,b,c,d should be answered.

58 1972-73 work history. Section A of 58 should be completed.

Military Service

64 Served in Armed Forces?

Information About the Past (Form B)

95 Ethnicity

Background Information

Names, addresses, and telephone numbers of respondent *and* his parents. Date of birth, sex, and social security number for Form B respondents. Signature permission for Form B respondents.

Immediately after receipt and check-in of a mail-returned questionnaire, it was routed to a manual pre-machine edit section where responses to the above key items were verified for presence and consistency. (Cost restraints and questionnaire complexity limited the editing to these items.) Failures were noted on an Edit Problem Sheet, which was routed with the questionnaire to the Telephone Tracing Department. The respondent was telephoned if possible so that inconsistencies could be clarified or corrected and missing data could be supplied.

Chapter IV

MAXIMIZING PARTICIPATION AND RESPONSE

The NLS population is young, highly mobile, and involved in decisions about future work and life patterns. Continued participation in NLS depends largely on success in keeping in touch with, motivating, and developing rapport with individuals in the survey population. Methods used to enhance the willingness and to increase the response included newsletters, thank-you letters, mailgrams, reminder post cards, and wallet-sized cards showing NLS involvement and serving as change-of-address cards. Several recommendations for maintaining cooperation and improving response were outgrowths of the first followup survey; these recommendations are given in section H of this chapter.

Response to the mail questionnaire will become more crucial and costly each year with each survey. For this first followup, the letter, post card, and telephone contacts were designed to produce a mail response rate of 60 percent, within resource and cost considerations. The remaining 40 percent were to be individually interviewed by Census Bureau field personnel.

A. Newsletters

In July 1973, RTI sent newsletters to base-year participants. Slightly modified newsletters were mailed to "resurvey" members in September 1973. The first newsletters reviewed the purpose of NLS, described student participation levels, announced that the first followup survey was underway, stated that a questionnaire would be mailed, stressed the importance of continued participation, and requested that the respondent return an enclosed card verifying current address or showing the needed corrections. Before the major mailout of the First Followup Questionnaire, resurvey members were mailed a revised version of the newsletter; it explained the study, elicited cooperation, and requested permission to use school record information to be collected from resurvey schools. Mailed in 1973 were 23,560 newsletters:

16,650 on July 27, 30, 31
400 on August 21
3,408 on September 26
3,102 on October 8, 9

Benefits of a newsletter were valuable leadtime in locating respondents for whom addresses had changed and being able to update the addresses prior to the questionnaire mailout.

B. Tracing Activities

About 90 percent of the more than 2,000 cases from the base-year and the resurvey efforts were traced by telephone (table 4-1). Mail was used for students, parents, and others who did not have telephones and for some colleges and State motor vehicle departments. Of the unsuccessful cases, 14 were sent to a local credit bureau; the bureau was unable to obtain current addresses for these.

In the Base-Year Student Questionnaire were requests for the student's name, address, social security number, and driver's license number; the name of the postsecondary educational institution the student planned to attend; and the names and addresses of two people who would always know where to get in touch with the respondent. Similar tracing information was requested in section F of the First Followup Questionnaire.

Table 4-1.–Telephone tracing cases, 1973-74

Reason for tracing	Cases				
	Traced	Completed		Unsuccessful	
		Number	Percent	Number	Percent
Total	2,067	1,890		177	
Newsletter undeliverable	1,565	1,493	95.4	72	4.6
Questionnaire undeliverable	502	397	79.1	105	20.9

Early in August of 1973, tracing to prepare for the first followup began as the postal service returned undeliverable newsletters; it continued through January 1974 as undeliverable questionnaires were returned. Information types and sources were:

1. Names, addresses, and telephones of parents, guardians, or relatives.
2. Names, addresses, and telephones of two people who would always know how to get in touch with the individual.
3. Name and location of postsecondary school the individual attended or planned to attend.
4. Name and location of the individual's employer.
5. Neighbors of the individual or his parents.
6. Principal or other contact at the secondary school attended.
7. State or registration and identification number of driver's license.
8. Local government agencies.
9. Armed Forces locator services or DOD rosters.
10. Institutional records (i.e., prison, police, mental).
11. Local credit bureau or a similar organization.

As new addresses were obtained from tracing activities, returned mail, or other sources, the computer file of names and addresses was updated.

C. Data Collection Activities

1. Mail-Return Responses

Questionnaires (Forms A and B) were mailed on 23-24 October 1973 to 22,654 students in the first followup sample. Events following this mailout (table 4-2) resulted in 13,928 returns and a 61.5 percent mail response rate by February 1, 1974. When necessary, additional questionnaires were mailed; 1,376 second questionnaires were mailed as a result of tracing activities from November 16, 1973 through January 4, 1974.

Not shown in the table are the 107 duplicates caused by timelag and error, 108 written refusals, a letter mailed with the second questionnaire to nonrespondents, a letter of thanks sent to all mail respondents, and the computer-generated, self-mailing ID and change-of-address cards sent to all participants after the first followup survey.

Figure 4-1 shows cumulative response rates to *all* mailouts. The bulk of the mail responses (61 percent) was received in the first 3 weeks after the October 1973 mailout. The data were 21 percent for the next 3-week period and 18 percent for the final period prior to February 1, 1974. Additional mail questionnaires were received later, during the field interview phase. The mail returns finally totaled 14,019 or 61.9 percent.

Table 4-2.--Overall mail returns and response rates, 1973-74

Event	Number mailed	Days elapsed*	Questionnaire response		
			Number returned	Cumulative total	Overall response rate (percent)
First questionnaire mailout	22,654	---	---	---	---
Reminder/thank-you post card	22,654	8	2,615	2,615	11.5
First prompting post card	14,053	23	5,277	7,892	34.8
Second questionnaire mailout	12,163**	37	1,723	9,615	42.4
Second prompting post card	11,945**	44	1,017	10,632	46.9
Mailgram requests***	10,313	53	1,502	12,134	53.5
---	---	72	1,423	13,577	59.8
---	---	102	371	13,928	61.5

*Days since the first questionnaire mailout of 10/23-24, 1973 (see fig. 4-1).

**On 11/27-28, post cards were mailed to 2,183 of the 12,163; on 12/4-5, second questionnaires were mailed to 3,080 of the 11,945. (Mailings were results of the special study, section IV of this report.)

***A final effort to increase the rate to 60 percent (a result of the special study).

Figure 4-1.--Cumulative response to all mailout types, 1973-74

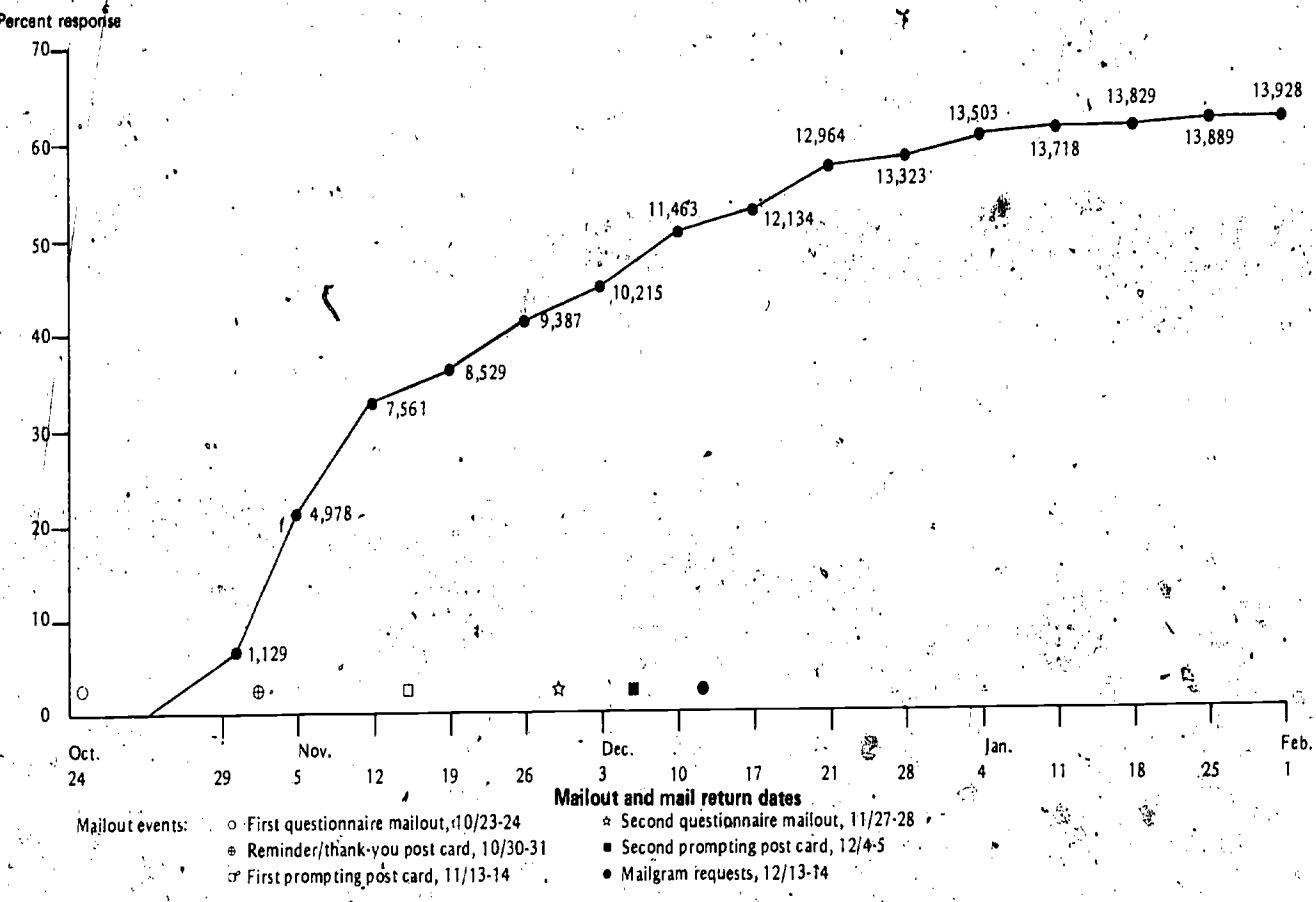


Table 4-3 compares mail response rates of the base-year/first followup participants by sex, race, region, urbanization, and family income as of the senior year in high school. Complete data were not available from base-year files on all variables for all students; for example, data on the resurvey respondents were missing sex, race, etc., which were gathered in the first followup in Form B. The response rates were highest for females (71.0 percent) and for Whites (70.9 percent). The South was lower (60.8 percent) than the other regions (average 66.8 percent). The degree of urbanization of the schools appeared to have little effect on response; the nine categories ranged from 61.4 to 68.9 percent. There was a spread of 19.6 percent in response rates affected by income; students of the higher income families generally responded at a high rate.

2. Effects of Mailouts

Although it was not possible to determine the effect of each mailout on the response rate, one conclusion was that no appreciable effect is felt until about 5 days after a mailout. This conclusion allowed

Table 4-3.–Mail response rates of subpopulations

Subpopulation group	Number of participants		Total response rate (percent)
	In sample	Responding	
Sex			
Male	8,074	4,864	60.3
Female	8,201	5,825	71.0
Race			
Black	2,023	989	48.9
White	12,287	8,711	70.9
Other	1,550	888	57.3
Region			
Northeast	3,742	2,522	67.4
Central	4,857	3,234	66.6
South	5,963	3,626	60.8
West	3,157	2,093	66.3
School urbanization			
1. 10 largest central city districts	1,915	1,225	64.0
2. Next 90 largest central districts	2,058	1,264	61.4
3. SMSA schools within 10 largest	1,599	1,102	68.9
4. SMSA schools within next 90 largest	2,834	1,897	66.9
5. All other SMSA schools	3,258	2,088	64.1
6. Outside SMSA, 75-100% urban	383	252	65.8
7. Outside SMSA, 50-74% urban	1,724	1,116	64.7
8. Outside SMSA, 25-49% urban	2,168	1,414	65.2
9. Outside SMSA, 0-24% urban	1,570	978	62.3
Income			
Less than $3,000	943	510	54.1
$ 3,000 to $ 6,000	1,621	965	59.5
$ 6,000 to $ 7,500	1,462	922	63.1
$ 7,500 to $ 9,000	1,297	830	64.0
$ 9,000 to $10,500	1,524	1,101	72.2
$10,500 to $12,000	1,216	831	68.3
$12,000 to $13,500	1,106	763	69.0
$13,500 to $15,000	837	615	73.5
$15,000 to $18,000	1,082	787	72.7
$18,000+	1,756	1,294	73.7

the estimates of response rates that are shown in table 4-4, but it did not allow for cumulative effects caused by multiple mail requests.

Table 4-4.–Estimates of effects of mailouts on response rates

Event	Overall response rate (percent)
Total	61
First questionnaire mailout	21
Reminder/thank-you post card	16
First prompting post card	8
Second questionnaire mailout	6
Second prompting post card	3
Mailgram requests	7

The overall mail response rate, slightly more than 61 percent (table 4-5), appears to be high–considering the questionnaire length (24-28 pages) and complexity and considering that no incentive was offered and that "resurvey" students (about 20 percent of the total) had had no prior NLS involvement. The response rates for the resurvey group lagged well behind those for the base-year sample. Resurvey individuals did not receive spring 1972 familiarization; the tracing information (only the addresses on file at the high school graduated from, more than a year before) was much less complete; NLS involvement was lacking; and the known mobility of this age group made all contacts difficult–all of these were detrimental to the overall response rate.

Table 4-5.–Response rates, by number of weeks after mailings

Weeks elapsed	Base-year participants rate	Resurvey participants rate	Difference	Overall response rate
	Percent			
2	28.0	13.6	14.4	24.8
3	36.9	19.2	17.7	33.0
4	42.3	23.4	18.9	38.1
5	46.5	27.1	19.4	42.2
6	50.1	29.8	20.3	45.6
7	55.3	34.9	20.4	50.8
8	59.1	38.9	20.2	54.6
10	64.0	43.5	20.5	58.4
15	65.2	47.1	18.1	61.5

3. Personal Interviews

Under a contract between OE and the Bureau of the Census, tracing information on 8,549 mail nonrespondents (appended to blank questionnaires) was turned over to Census workers early in January 1974. This had been preceded by preliminary December listings of nonrespondents distributed by zip code and Census district so that the districts could plan workloads.

Although the purpose was to locate and obtain personal interviews with all nonrespondents, a restriction of 50 miles had to be imposed (due to the emerging fuel crisis) beyond which a telephone interview would have to suffice. This telephone interview contained a subset of 22 key questions–those deemed most critical for analyses–because the entire questionnaire was too long for effective telephone interview.

Fieldwork began in late January 1974. By the end of March all interviews (personal and telephone) had been completed for all available sample members. Of the 8,549 cases referred, 7,331 were interviewed, for a gross completion rate of the interview subsample of 85.8 percent. For personal interview, administration time averaged 45 minutes; by telephone it was about 15. Just over 10 percent of these cases were completed by telephone.

4. Checks on Data Quality

In November 1973, editors were recruited and trained to perform data quality checks as the first followup returns came in. Each returned survey instrument was scan-edited page by page to detect critical omissions and inconsistencies. The scan-edit averaged 15 minutes per questionnaire. This review was designed to verify–

- That in general, the respondent had understood and followed the instructions (i.e., he had not given obviously false or capricious responses or in some other way shown an insincere effort in completing the questionnaire);
- That the major critical skip patterns were correctly interpreted and that no block of applicable items was omitted;
- That each instrument was properly identified and all items required for tracing were completed;* and

*This verification was used in the check-in process and in the data quality checks because the critical nature of these items justified the added cost.

That questions deemed critical for minimal analyses relevant to survey objectives were answered adequately.

Critical items selected for review included up to 22 questions, depending on skip patterns; these covered:

- Current activities and plans,
- Education and training since high school,
- Civilian work experience since high school,
- Military service, and
- Background (tracing) information.

The list of key items is in chapter III. Only key questions were examined for clarification or correction by the respondent. No other information was checked.

The goals were not only to obtain answers to key questions but also to ensure that the answers were consistent among themselves. Tables 4-6 and 4-7 show the results of the premachine manual edit for mail and Census Bureau completions.

If a questionnaire failed the edit checks, its problems were noted on an Edit Problem Sheet, the sheet and the questionnaire were routed to the Tracing Department, and the staff (also trained editors and interviewers) telephoned the respondents. A total of 28.9 percent or 4,022 mail questionnaires failed to meet the minimum requirements established and were turned over to the Tracing Department for telephone followup; of these, 3,822, or 95.0 percent, were completed successfully. Of the 34 Census Bureau questionnaires that failed edit specifications, 31 (91.2 percent) were successfully completed.

Table 4-6.--Telephone followups of mail-return questionnaires that failed edit specifications

Response category	Mail-return questionnaire*					
	Form A		Form B		Total	
	Number	Percent	Number	Percent	Number	Percent
Total	2,982	100.0	1,040	100.0	4,022	100.0
All items completed	2,828	94.8	947	91.1	3,775	93.9
One or more key items not completed	22	0.7	25	2.4	47	1.2
Refusals	8	0.3	4	0.4	12	0.3
Respondents in military overseas	25	0.8	9	0.9	34	0.8
Respondents without telephone	53	1.8	26	2.5	79	2.0
Respondents with unlisted numbers	8	0.3	3	0.3	11	0.3
Other	38	1.3	26	2.5	64	1.5

*Clarifications of Forms A and B of the First Followup Questionnaire continued through April 29, 1974.

Table 4-7.--Telephone followups of Census questionnaires that failed edit specifications

Response category	Bureau of the Census questionnaire*					
	Form A		Form B		Total	
	Number	Percent	Number	Percent	Number	Percent
Total	25	100.0	9	100.0	34	100.0
All items compelted	24	96.0	7	77.8	31	91.2
One or more key items not completed	0	0.0	1	11.1	1	2.9
Refusals	1	4.0	0	0.0	1	2.9
Respondents without telephones	0	0.0	1	11.1	1	2.9

*Clarifications of these questionnaires continued through April 29, 1974.

D. Telephone Survey

A telephone interview was conducted for a small subsample (N = 393) of those who had not responded by mail as of December 17, 1973. The primary purpose was to find out some reasons for mail nonresponses. The survey was conducted by the Telephone Tracing Department during the period December 18-28, 1973 with 298 base-year and 95 resurvey nonrespondents selected by random sampling. There were 220 (56 percent) sample members contacted, of whom 169 were base-year participants and 51 were resurvey members. Also, 98 family or household members were contacted (83 base-year). Reasons for nonresponse varied widely, including 19 who said they "never got the questionnaire," 26 who "lost the questionnaire," and 49 who "haven't had time."

A secondary purpose of this telephone survey was to probe for actions to improve the response rate. Only 90 persons answered in this regard and of the 540 total responses given (90 respondents x 6 answers each), only 33 percent were positive about the various suggestions (e.g., "receiving more information about the study," "receiving a $5 incentive").

E. Special Study

Knowing that NLS would continue beyond the first followup, a plan was developed to obtain information for improving strategies for subsequent efforts without disrupting the ongoing high-response-rate efforts. Figure 4-2, the design of the followup procedure just discussed, indicates special supplementary samples and activities. It was not developed as a rigorous design to be used for comparisons of experimental treatments, but more as a practical guide for future NLS work.

Broadly speaking, this special study was designed to compare (1) air mail and certified mail questionnaires, (2) telephone calls prior to the respondents' receipt of the questionnaires and prompting calls made after receipt, and (3) mailgrams and prompting post cards. Table 4-8 indicates for each group shown in figure 4-2 the date, the action, and the subgroup response rates as of January 4, 1974. All of the 999 sample members who were sent certified mail questionnaires were base-year participants. Even correcting for resurvey members among the other two initial groups, certified mail responses were about 10 percent over air mail responses. Telephone contact had the highest impact on response rate. It was also by far the most expensive. Based on the costs and the need to reach a large number of nonrespondents in a short time, mailgrams appeared to be the most cost effective.

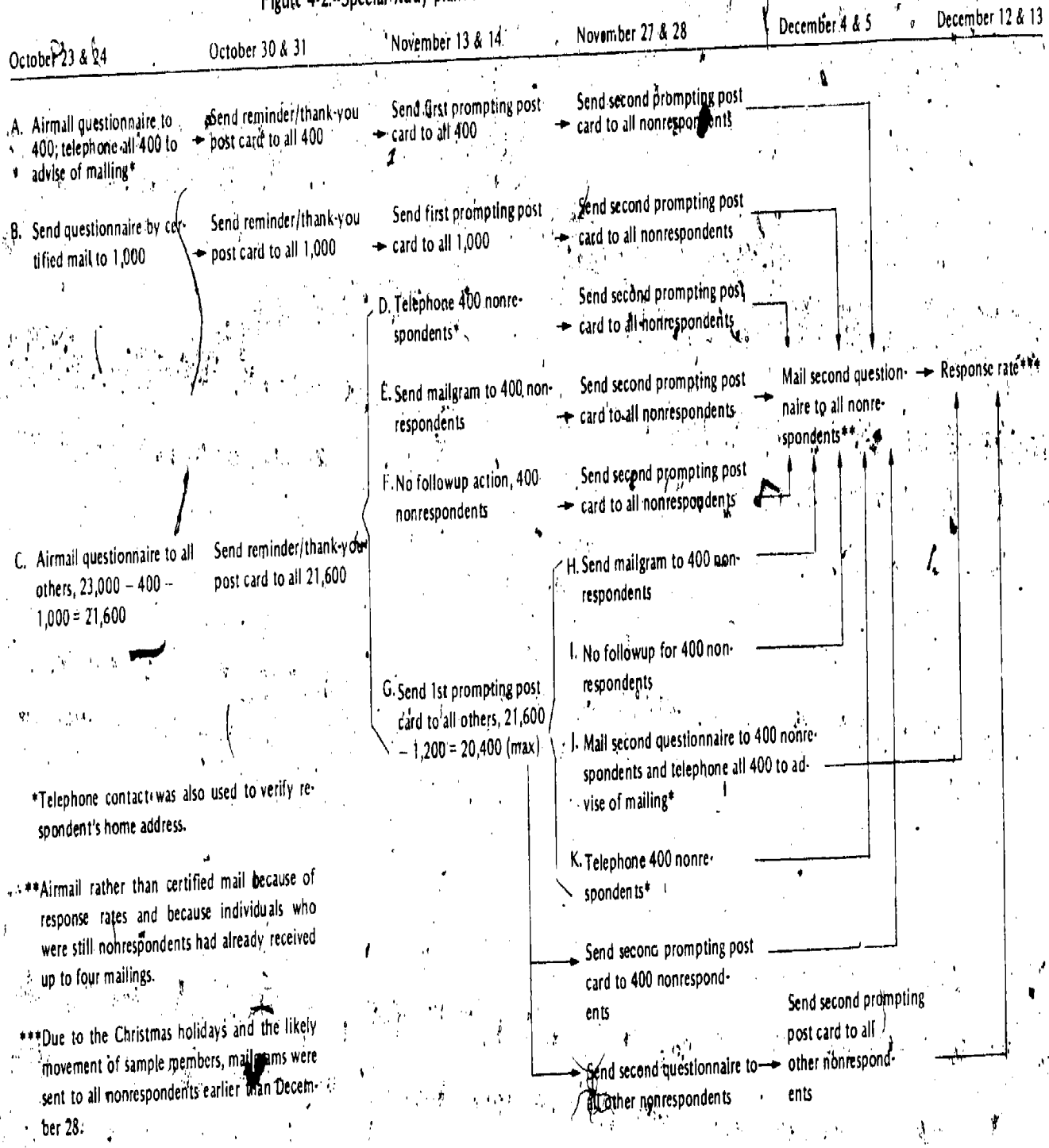
Once the response rates (table 4-8) for procedures, and their costs and scheduling requirements were known, a decision was made in early December 1973 to send mailgrams to all nonrespondents; based on data at hand, the final response rate with this procedure was expected to be close to 60 percent. The 10,313 mailgrams sent on December 13 and 14 had an even more positive effect than anticipated. The overall response rate was 61 percent. Admittedly, it is hard to know how many without this effort might have come in from the college students who had spare time during the holiday season or who returned home for vacation and found the questionnaire.

Table 4-8.--Special study mail-return response rates as of January 4, 1974

Date and event*		1973-74 Number mailed	Number contacted	Number returned	Subgroup response rate (Percent)
10/23	A. Telephone/air mail	404	318	273	85.8
	B. Certified mail	999	913	690	75.6
	C. Air mail	21,251	20,762	13,170	62.4
11/13	D. Telephone	400	365	251	68.8
	E. Mailgram	398	386	198	51.3
	F. No followup	400	400**	184	46.0
11/27	H. Mailgram	400	394	153	38.8
	I. No followup	400	400**	116	29.0
	J. Telephone/mail	399	330	183	55.5
	K. Telephone	400	350	198	56.6

*See Figure 4-2.
**Estimated.

Figure 4-2.--Special study plan: NLS mailout and followup procedures, 1973-74



*Telephone contact was also used to verify respondent's home address.

**Airmail rather than certified mail because of response rates and because individuals who were still nonrespondents had already received up to four mailings.

***Due to the Christmas holidays and the likely movement of sample members, mailgrams were sent to all nonrespondents earlier than December 28.

F. Resurvey School Data Collection

During the week of March 18, 1974, a packet was sent to resurvey schools. In it were a computer listing of the resurvey students selected to participate, two copies of the School Questionnaire, individually labeled SRIF's for each sample member, instructions for completing these items, a postpaid return post card for acknowledging receipt of these materials, and a postpaid return envelope for mailing completed materials. No Counselor Questionnaires were requested. As in the base-year survey, participating schools were paid $40 honararia to cover the cost of data collections.

During the following week, post cards acknowledging receipt of packets began arriving and survey specialists began receiving telephone calls from resurvey schools (about 100) regarding problems—namely, confusion over ID numbers, the need for maiden names, misunderstandings of the SRIF's, and permission codes and signature verifications for permission. By the April 26 deadline for receipt of data collection materials, 155 acknowledging post cards and 85 completed packets had been received. Prompting telephone calls were made to the schools from whom post cards, but no data collection packets, had been received. No responses had been received from 105 schools; these 105 received reminder telephone calls from the project director.

About 23 May, mailgrams and letters were sent to 1,344 resurvey members who had neglected to sign the permission line for acquiring SRIF data. In the fall of 1974, a final effort was made to secure the remaining SRIF's—the resurvey schools were asked to submit transcripts only for remaining resurvey students, and the data was transfered to SRIF's.

G. First Followup Data Collection Summary

The first step in data collection for the first followup involved an extensive tracing operation to update name and address files received from the previous contractor for the base-year survey. A total of 18,672 of these individuals from 1,044 schools were used for the first followup survey. In addition, individuals from 256 resurvey schools were added to the base-year lists to give a first followup survey sample of 1,300 schools and 22,654 individuals. Newsletters were developed and mailed not only to encourage participation but also to serve as vehicles for updating names and addresses. When mail was returned by the postal service as undeliverable, telephone tracing procedures were used to obtain current addresses where possible.

Questionnaires were mailed to the last known addresses of the sample of 22,654 on October 23-24, 1973. This was followed by a planned sequence of reminder post cards, additional questionnaire mailings, and reminder mailgrams to nonrespondents. Active mail-return efforts continued through December 1973; by early February 1974, the questionnaire-return rate by mail was 61.5 percent. Questionnaires continued to arrive through the mail during the field interview phase of the survey.

The names and addresses of those sample members who failed to mail back their questionnaires were then turned over to the Bureau of the Census for personal interview in accordance with a Bureau arrangement with OE. This personal interview phase of first followup data collection continued until April 1974, at which time the overall response rate had been increased to 94.2 percent (21,350 respondents out of 22,654).

H. Recommendations

Recommendations based on the special study of the first followup survey are for improvements in the questionnaire, the prequestionnaire letter, the newsletter, the ID card, and the mailout plan.

1. Simplify the Questionnaire

Directions and procedures for answering questions and following skip patterns need to be improved to elicit correct answers to appropriate questions. An example is needed on the inside of the front cover to show how to answer questions.

2. Prepare an Advisory Letter

Several weeks before the questionnaire mailout, an advisory letter should be sent to parents of sample members whose first followup address is the same as that of the parents. This letter would explain that a package of survey materials for their son or daughter (a 1972 senior) will arrive within a few weeks, and it would ask the parents to either forward the package (instead of holding for the return from college, etc.) or to call collect with another address for a second mailing.

Approximately 1 week before the questionnaire mailout, a brief advisory letter should be mailed to all sample members at their most likely address at the time of survey to explain that they should be receiving survey materials within 10 days by certified mail. Certified mail is one way to imply importance and to elicit responses at minimal cost.

3. Use Mailgrams and Prompting Post Cards

Half of the nonrespondents to the first questionnaire mailout of the second followup should be sent mailgrams; the other half, prompting post cards. For the second questionnaire mailout, this should be reversed: those who received mailgrams before should receive post cards, and vice versa. Total responses to these two should help determine the cost effectiveness of mailgrams.

4. Improve the Newsletter and ID Cards

The newsletter should contain more results-oriented data to elicit interest in the growth and activities of the class or age group. Adoption of a newspaper format and inclusion of personal interest stories about sample members, staff, and processes should be helpful.

Several hundred ID cards were returned in 1973-74 with correct addresses; however, the respondent's name and address were easily smudged and became illegible. A new, noncarbon format should be adopted.

5. Consider Other Tracing Aids

Time lapses between data collections (15 months elapsed between base year 1972 and the first 1973 mailout of newsletters) can decrease the efficiency of tracing efforts. Therefore, new tracing aids should be considered for improving followups. Even though the use of the credit bureau system was not helpful in the first followup, it may be that as study members mature and establish their own households it can be a useful, low-cost source of tracing information.

Chapter V

PREPARATION OF THE NLS DATA FILE

Questionnaires returned by mail, either from individual sample members or from Bureau of the Census interviewers, were routed on a flow basis to a central check-in point. First, each respondent's ID number and batch number were transmitted to the data processing section for a daily count of the questionnaires received. Then, batches were routed to the premachine (manual) editing section to see if each questionnaire contained the minimum set of key data. Questionnaires which passed this check were transmitted to the direct data entry section to be transformed into machine readable form. For questionnaires which failed the check, telephone contacts were made with the respondents to resolve problems uncovered in manual edit; after resolution, the questionnaires were transmitted to the direct data entry section for encoding. At each point along the route, events were entered into the automated survey support system so there would be constant monitoring of the location of all questionnaires in the document flow process.

After data were encoded, questionnaires were placed in temporary secure storage prior to microfilming and eventual destruction. (Microfilming of the first followup data was completed in the summer of 1974.) The microfilming was initiated to provide greater assurance of confidentiality, to simplify accessibility of original data, to reduce storage costs (several tons of hard copy were transferred to microfiche), and to ensure against loss by separate storage of a second filmed copy.

A. Pre-Machine Editing

The pre-machine editing centered around the following ideas: If "key" questions were answered or could be presumed by responses to related items, a questionnaire would be accepted—no matter how little other information was given; however, if any key data (described in chapter III) were missing, the respondent would be telephoned to complete the record.

After a questionnaire had passed the editing requirements, the alphabetic data in a few parts of the questionnaire were manually coded into numeric form before transformation into machine readable form. Two questions (49a,55a) about job area or occupation for 1972 and 1973 were transformed into the corresponding 3-digit codes specified in the Census Occupational Classification System; the parental occupation question (79) was likewise coded.* Postsecondary school identification (college, university, vocational/technical) was transformed into six-digit FICE** codes by using a master index provided by OE. For fields of study (22b,26a,36a,60b), both the four-digit academic subdivisions provided by the HEGIS (Higher Education General Information Survey) taxonomy and the six-digit HEGIS technological and occupational schemes were used.† Finally, the responses to question 43b were coded numerically to indicate type of license, certificate, or diploma.††

There were 18 questions in the First Followup Questionnaire for which respondents could select either fixed-choice (closed-response) answers or write in answers for the "other" option. In every case possible, the "other" option was reclassified by the manual editors into the fixed-choice options; when reclassification was not obvious or logical, the "other" was retained in the coding, but the alphabetic description was not included on the data file. There were four questions (44ba,69b,69c,70b) in the First Followup Questionnaire that were not numerically coded (i.e., they were coded as written on the questionnaire).

NLS first followup data were transformed to machine readable form by using direct data entry pro-

*Codes are in appendix C of the *Base-Year and First Follow-up Data File Users' Manual*. This document is hereafter referred to as the *Data File Users' Manual* or the *Users' Manual*.

**FICE codes are six-digit serial numbers used to identify U.S. institutions of higher education (those offering 2 or more years of college); these codes are available in NCES directories and from the OE computerized Vendor's File.

†Appendix D-1, *Users' Manual*.

††Appendix D-2, *Users' Manual*.

gramable terminals. Major advantages of this procedure were higher speed, fewer processing steps, and lower transcription error rate. The overall error rate was less than 0.3 percent. The terminals were programed to accept a specific range of values for most of the data and specific field widths for all data. Response ranges for fixed-format variables decided upon prior to data entry were programed for the terminals; other responses were coded to indicate they were outside of the ranges.

B. Error Coding

A set of "error" codes was applied uniformly across the file to indicate classes of erroneous or missing data. The codes are explained below:

93–*Partial response.* Used for questions with the two-column response format to indicate whether each subitem applies or not. If at least one was answered, the unanswered subitems were coded 93.§

93–*Don't know.* (self-explanatory).

95–*Out-of-range response.* Used when a response or transcription exceeded an acceptable range.

96–*Multiple response.* Used when there were several answers to a question when the directions call for only one.

97–*Refusal.*

98–*Blank, or nonresponse.* Used for nonresponse cases not identified as legitimate nonresponse (see section C-3 below).

99–*Legitimate nonresponse.* Used when the respondent should not and did not answer the question (i.e., he was routed around it) or did not answer an entire instrument. In the latter case, all fields were coded 99 (see section C-3 below).

The above error codes were applied to items with two-digit fields only; these are by far the most common type in the file. For items with three-digit fields, the error codes have two leading 9's followed by the digit indicating the error class that is, 993 through 999. For four-, five-, and six-digit fields, the number of leading 9's was increased accordingly.

§First followup questions that received 93 codes were 1, 16, 22a, 24, 29b, 31, 35, 38, 44b, 48b, 54b, 62, 63, 71, 76.

C. Machine Editing

Preparing the NLS data tapes for public release required "hard copy" (source document) spot checks and machine editing to recode all uninterpretable responses and some logical recoding of the responses. Thus, the final data file contains only valid response codes, erroneous or missing data codes, and "logically recoded values" with indicators for the reason for recoding (section 3 below).

Three machine-editing programs—range checks, consistency checks, and routing checks—were used sequentially for the first followup data. These programs were not applied to the NLS base-year data; these data, edited by the previous contractor, were either reformated or recoded to achieve consistency with the first followup data on file.

1. Range Checks

The first program checked the responses to each fixed-choice item against a range of acceptable values and "flagged" and recoded with 95 any value outside of the range. Acceptable ranges for 72 numerically coded open-response items and the number of responses outside these ranges are tabulated in table 5-1. In almost all cases, some responses were outside of these ranges. In general, these responses were logically possible but highly improbable. It was felt that some outlying responses could provide additional data and that it was best to leave them in the field to provide as faithful a transcription of the original records as possible.*

2. Consistency Checks

The second program checked the consistency of an individual's responses over the entire questionnaire against a set of 94 internal checks (or response comparisons) selected *a priori*. The program read the responses and flagged those which were failed; then an index was computed for each record to indicate the internal consistency and to provide the user a rough indication of data quality.

*Computer tapes are available with the out-of-range values replaced by the appropriate 95 out-of-range code.

Table 5-1.–Out-of-range responses to free-response questions calling for numeric answers

Open-response question number	Range of acceptable values	Number of out-of-range responses	Open-response question number	Range of acceptable values	Number of out-of-range responses
3A	1-12	6	47BB	0-9,000	1
3B	71-73	15	47CB	0-9,000	0
7B	1-12	6	47DB	0-9,000	0
7C	67-74	2	47EB	0-9,000	0
8B	1-4	1	47FB	0-9,000	0
11A	0-20,000	39	47GB	0-9,000	0
11B	0-20,000	16	49FA	1-12	5
11C	0-20,000	28	49FB	63-74	26
11D	0-20,000	6	50A	1-50	684
11E	0-6,000	7	50B	0-1,000	83
11F	0-6,000	1	55FA	1-12	7
11G	0-10,000	7	55FB	63-72	164
11H	0-10,000	0	55GB	1-12	27
27AA	1-12	4	55GC	72-74	41
27AB	68-73	9	56A	1-50	539
27C	0-50	38	56B	0-1,000	75
33AA	1-12	2	58A	0-52	51
33AB	68-72	125	58B	0-52	2
33C	0-50	57	58C	0-92	0
40CA	1-12	1	66B	1-12	3
40CB	68-74	4	66C	68-74	3
40DB	1-12	2	66D	1-12	2
40DC	68-74	1	66E	69-72	7
41CA	0-92	0	67A	1-12	4
41CB	0-92	0	67B	72-74	10
41CC	0-120	21	74B	1-12	6
46AA	0-9,000	1	74C	68-74	23
46AB	0-12	176	82DA	0-9,000	0
46BA	0-5,000	21	82DB	0-9,000	0
46BB	0-2,000	65	82DC	0-9,000	0
46BC	0-1,000	47	83DA	0-9,000	0
46BD	0-1,000	73	83DB	0-9,000	0
46BE	0-1,000	115	83DC	0-9,000	0
46BF	0-9,993	0	84DA	0-9,000	2
46BG	0-4,000	29	84DB	0-9,000	0
47AB	0-9,000	3	84DC	0-9,000	1

3. Routing Checks

The First Followup Questionnaire contains 33 questions that either implicitly or explicitly directed a respondent around those questions that did not apply. To determine if the respondent correctly followed the routing patterns, a routing-check program was developed and implemented. It read each record and flagged responses that were inconsistent with the subsequent pattern of response. The flags indicated the type of inconsistency detected:

a. Inconsistency occurring when the response to a routing item indicates that the questions within the routing pattern should have been skipped but were not. These were recoded by adding 20 to the original response code.

b. Inconsistency occurring when the response to the routing item indicates that the questions should have been answered but were not. Recoding added 40 to the original response code.

c. Inconsistency occurring as a combination of the first two types when the response to the routing item indicates that certain questions which should have been skipped were not (type a) and others should have been answered and were not

(type b). Recoding added 60 (20 + 40) to the original response codes.

Examples in the *Data File Users' Manual* clarify the testing and recoding procedures employed by the routing-check edit program, and appendix E.1 of the manual lists codes for consistent and inconsistent responses to the 33 routing questions in the questionnaire.

The routing-check program also differentiated between legitimate nonresponse (code 99) and illegitimate nonresponse (code 98). Legitimate nonresponse pertains to questions that the respondent was routed around. If a respondent was routed into a block, any nonresponse to those items is illegitimate; if the routing pattern is answered inconsistently (20, 40, or 60 added to it) with the routing instructions, any nonresponse to those items is illegitimate. The only time that a nonresponse was coded 99 was when there was an unflagged response to the question that routed the respondent around a group of questions. If a response pattern did not clearly indicate which questions the respondent should have answered, the nonresponse was coded 98. In some of the more complex routing patterns, nonresponse was coded 98 for a large section of items due to one inconsistency. The user should be careful in interpreting the 98 and 99 codes to avoid overestimating the number of illegitimate responses.

For some analyses, the user may wish to redefine legitimate and illegitimate skips. Appendix E.2 of the *Data File Users' Manual* lists all the questions in the first followup that should be checked because they are inside of one or more routing patterns. A question not in this listing is not in any routing patterns. To determine the legitimacy of nonresponse, all of the routing questions that cover a question should be considered.

D. The NLS Data File Indices

Two kinds of composite indices—quality and analytic—were developed and placed in the NLS data file. The quality indices quantify the amount and quality of data in the individual record. The analytic indices (derived from global considerations of the entire file) are classification variables used to group the individual records.

1. Quality Indices

Four quality indices were developed to quantify the amount and quality of First Followup Questionnaire data in each record. Each index was allocated a three-digit field on the record. Table 5-2 presents the quality indices.

Consistency index (CS) represents the percentage (truncated set of 94 checks) failed by an individual. The index was computed as

$$CS = \frac{\sum_{i=1,94} X_i}{94} \cdot 100$$

where X_i is 0 if respondent passed or 1 if failed check i.

Out-of-range index (OR) represents the percentage of out-of-range responses for an individual. It was computed as

$$OR = N/D \cdot 100$$

where N is the number of items coded 95 (out-of-range) and D is the number with response other than 99 (legitimate skip).

Routing index (RI) represents the percentage of routing questions ambiguously answered (i.e., unanswered or answered inconsistently with the subsequent response pattern). The index was computed as

$$RI = N/D \cdot 100$$

where N is the number of flagged routing questions and D is the number of routing questions with responses other than 99.

Completeness index (CIA through CIE) represents the percentage of items with responses (not coded 93-98) for each of five sections (A-E) in the questionnaire. The index was computed as follows

$$CI_x = N/D \cdot 100$$

where x is the section (A,B,C,D, or E), N is the number of items with valid responses, and D is the number of items with responses other than 99.

Table 5-2 indicates lower completion rates for sections A and B of the questionnaire than for other

Table 5-2.--Quality of individual responses to First Followup Questionnaire

Quality index	Data file code	Questionnaire responses: Number	Percent
1. Consistency index	0*	9,574	44.84
	1	5,781	27.07
	2	3,204	15.00
	3	1,321	6.19
	4	670	3.14
	5	323	1.51
	6-100	477	2.23
2. Out-of-range index	0**	19,018	89.10
	1	2,066	9.67
	2	145	0.68
	3	50	0.23
	4	33	0.15
	5	6	0.03
	6-100	32	0.15
3. Routing index	0+	12,187	57.08
	3-5	4,403	20.62
	6-10	3,147	14.74
	11-15	1,024	4.80
	16-25	473	2.22
	26-35	85	0.40
	36-100	31	0.15

4. Completion index	Questionnaire sections: A		B		C		D		E	
(percent)	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
100-91++	4,074	19.08	6,854	32.11	16,494	77.26	20,394	95.53	17,077	79.99
90-81	10,203	47.79	6,232	29.19	920	4.31	86	0.41	913	4.28
80-71	3,558	16.67	1,907	8.94	920	4.31	79	0.37	695	3.26
70-61	1,949	9.13	1,281	6.00	581	2.73	36	0.17	422	1.98
60-51	338	1.60	964	4.52	308	1.45	20	0.10	267	1.25
50-41	84	0.40	710	3.33	449	2.11	369	1.73	456	2.14
40-0	1,144	5.36	3,402	15.94	1,678	7.86	366	1.72	1,520	7.12

*0 means no consistency checks failed.
**0 means no out-of-range numeric responses.
+0 means no routing questions answered ambiguously.
++100 percent means all items completed.

sections. This result is artifactual and may be misleading. The major utility of this index is in judging the completeness of an individual's responses within a section, not between sections, because the sections vary in complexity, number of routing patterns, and types of information elicited. Section A contains fewer routing patterns than other sections, but contains some items (e.g., 1G, 16F) for which one may expect larger nonresponse; also the routing check rules for item 1-1 specify code 98 for all nonresponses and thus inflate the nonresponse. Review of the indices *within* a section will illustrate the typical completion rate and identify outlying individuals.

The utility of quality indices is in judging the credibility of individual records. They are of no use in judging data over respondents such as item responses, since the real test of item response quality is the over-subjects distribution. It is recommended that these indices not be used for discarding subjects unless one's concern is with either the entire instrument or subsection.

The machine editing indices just discussed were

modified somewhat in developing a second level of edit version of the NLS data file—deliverable as a part of second followup survey work. Users of the second level of edit file should consult the *Data File Users' Manual* for that file.

2. Analytic Indices

Two composite indices—an ability index and a socioeconomic status (SES) index—involve several components each and required several steps during derivation. Since other components and procedures may be used in deriving such indices, users are encouraged to decide whether the indices included here are appropriate for their needs.

Ability Index. Each NLS participant's code of 1, 2, or 3 indicates an ability composite score in the lower, middle two, or upper quartile:

3 = upper quartile if K 225.7497,
2 = middle two quartiles if ⩽ 225.7497 and ⩾ 181.5461,
1 = lower quartile if > 181.5461.

These were derived from vocabulary, reading, letter group, and mathematics test scores. A simple sum over the four scores (each with a mean of 50 and a standard deviation of 10) was accepted as the general ability index because an analysis yielded a first principal component with essentially equal weights for the four and a rotated factor which explained the most common variance gave essentially the same result.

SES Index. Each NLS participant was also assigned 1, 2, or 3 indicating an SES composite score in the lower, middle two, or upper quartile:

3 = upper quartile if > +0.4735,
2 = middle quartiles if ⩽ +0.4735, and ⩾ −0.4735,
1 = lower quartile if < −0.4735.

The composite has five components: father's education, mother's education, parent's income, father's occupation, and household items. Factor analysis of the five revealed a common factor with approximately equal weights for each. Then each was standardized. An equally weighted combination of the five standard scores yielded the SES composite scores.

E. The NLS Data Files

Two data files were developed and derived: the NLS master file and the NLS release file. Each of them has a computer-generated NLS variable list and NLS response list to provide details of the data stored in each record of the file. The variable list contains the name and description of each variable, the field or character positions containing each variable, and a response list reference code for each variable. The response list catalogs the valid response codes for types of variables in the variable list.

The variable lists reveal that a number of the NLS variables (items) from the base-year and first followup were deleted from the master file to create the release file, primarily to protect the confidentiality of individual data. Others were deleted or modified due to excessive prior editing or poor response. The complete computer data file, or master file, for which general release was not contemplated, contains all data collected.

The variable list shows that the data in each record are, *in general*, grouped and ordered by identification codes, data indicators, Test Book data, SRIF data, Base-Year Student Questionnaire data, First Followup Questionnaire data, quality indices, sampling weights, school data, analytic indices, and activity statuses.

The number of respondents and amounts of data available for subpopulations and for combinations of instruments for two survey periods are tabulated in Table 5-3. For more details of the release file, consult appendix F of the *Data File Users' Manual*.

Table 5-3.--Number of responses by subpopulations to survey instruments

Population group	Survey instrument							
	B-Y (Base Year Student Questionnaire)	FFU (First Follow-up Questionnaire)	SRIF (School Record Information Form)	TB, (Test Book)	B-Y and FFU	B-Y, FFU and SRIF	B-Y, FFU, and TB	B-Y, FFU, SRIF, and TB
Total	16,683	21,350	20,651	15,859	15,635	15,618	14,863	14,848
Sex								
Male	*8,275*	*10,463*	10,233	7,894	7,665	7,658	7,307	7,301
Female	8,397	10,841	10,376	7,953	7,967	7,957	7,553	7,544
Unclassifiable	13	46	42	12	3	3	3	3
Race								
White	12,656	15,272	14,721	12,111	11,949	11,938	11,442	11,433
Black	2,083	2,739	2,592	1,906	1,920	1,917	1,766	1,763
Other	1,605	1,829	1,816	1,508	1,468	1,465	1,379	1,376
Unclassifiable	339	1,510	1,522	334	298	298	276	276
High school program								
Academic	6,811	8,511	8,312	6,531	6,468	6,465	6,204	6,201
General	5,665	7,492	7,253	5,363	5,235	5,224	4,955	4,944
Voc/tech	4,201	5,148	5,063	3,956	3,927	3,924	3,699	3,698
Unclassifiable	6	199	23	9	5	5	5	5
Region								
North	3,618	4,483	4,316	3,521	3,420	3,420	3,323	3,323
Central	4,568	5,541	5,468	4,122	4,292	4,288	3,875	3,873
South	5,513	7,691	7,242	5,382	5,186	5,178	5,057	5,049
West	2,984	3,635	3,625	2,834	2,737	2,732	2,608	2,603
Unclassifiable	---	---	---	---	---	---	---	---
Ability								
Low	4,788	4,392	4,783	4,798	4,382	4,374	4,382	4,374
Medium	7,000	6,600	6,997	7,008	6,592	6,585	6,592	6,585
High	4,052	3,890	4,052	4,053	3,889	3,889	3,889	3,889
Unclassifiable	843	6,468	4,819	---	772	770	---	---
SES								
Low	5,076	6,423	6,227	4,775	4,735	4,729	4,458	4,453
Medium	7,816	9,635	9,393	7,448	7,320	7,310	6,971	6,962
High	3,667	4,686	4,499	3,525	3,506	3,505	3,370	3,369
Unclassifiable	124	606	532	111	74	74	64	64

Chapter VI

WEIGHT CALCULATIONS

The sample for the NLS study is highly stratified, multistage, and clustered. As a consequence of the complex design, each observation (response) must be weighted in order to obtain unbiased sample estimates of population parameters. For all students sampled, the unadjusted weights are calculated as the inverses of the probabilities of being included in the sample. For several sets of nonrespondents, adjusted weights were calculated using the weighting-class methods described herein.

A. Unadjusted Student Weights

First, it was necessary to determine which schools and students were "in sample" for the 1972 NLS project. The NLS sample design included 1,200 primary sample schools (2 per final stratum) and 21,600 students (18 per school). The number of schools was increased (up to three or four per stratum) by using backup schools in the base-year and first followup surveys and by obtaining responses from all primary sample schools in the resurvey. Included in the NLS sample were--

1,153 participating primary sample schools
21 nonparticipating (no 1972 seniors) primary sample schools
131 participating backup sample schools
18 extra base-year backup sample schools
16 augmentation sample schools

1,339 NLS sample schools

The release tape contains data for students representing 1,318 schools--all of the 1,339 schools in the final NLS sample except the 21 primary sample schools with no 1972 seniors.

It would have been possible to reduce the number to the 1,200 originally intended, but this alternative was not chosen since considerable base-year student data would have had to have been discarded. The alternative chosen was to include all sample schools with students who had completed either a Base-Year Student Questionnaire or a First Followup Questionnaire.

1. School Selection Probabilities

For each school in the original 600 final strata, the sample inclusion probability, P_{hi}, was calculated as

$$P_{hi} = n_h\left(\frac{A_{hi}}{A_h}\right)$$

where n_h = number of schools in the final NLS sample for stratum-h,
A_{hi} = size measure for school-i of stratum-h, and
A_h = sum of size measures for all schools in stratum-h.

For schools in the smallest size strata (those with < 300 seniors), the size measure was either the estimated number of seniors in the school or 18, whichever was larger. For schools in the two largest size strata, the size measure was equal to 1. Thus, schools in the smallest size strata were selected with probabilities proportional to the number of seniors, and schools in the two largest size strata were selected with equal probabilities.

A sample of 200 school districts was canvassed during the base year to identify public schools which were not in the sampling frame. A total of 45 schools were identified, and 23 were randomly selected as an augmentation sample.[1] School selection probabilities, P_i, were calculated for each of the 45 schools, as suggested by the NLS sampling contractor.[2] The 16

[1] Educational Testing Service. 1973. *The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972--Final Report*. Princeton New Jersey.

[2] WESTAT, Inc. 1972. *Sample Design for the Selection of a Sample of Schools With Twelfth-Graders for a Longitudinal Study*. Rockville, Maryland.

participating augmentation schools were grouped into strata (h = 601, 602, ..., 608). For these 16 schools, the sample inclusion probabilities, adjusted for school nonresponse, were calculated as

$$P_{hi} = \left(\frac{16}{23}\right)\left(\frac{1}{2}\right) P_i .$$

2. Student Weights

The number of students from each sample school was taken as the number of eligible students offered a chance to be in the sample. This included all sample eligibles (respondents and nonrespondents), but excluded ineligibles such as dropouts, early graduates, and adult education students.

There were 23,451 sample students, of whom 16,683 completed a Base-Year Student Questionnaire and 21,350 completed a First Followup Questionnaire. For each of the 23,451 selected students, the unadjusted student weight, W_{uhij}, was calculated as

$$W_{uhij} = \frac{1}{P_{hi}} \cdot \frac{N_{hi}}{n_{hi}}$$

where N_{hi} = number of senior students in school-hi, and
n_{hi} = number of sampled students in school-hi.

The sum of the unadjusted student weights is an estimate of the total number of 1972 high school seniors in the population. If all of the selected students had completed the survey instruments, these weights would be appropriate for the analyses of student data.

B. Nonresponse Adjustments

Handling nonresponse when analyzing survey data is a problem. In general, the mean values of most survey variables are different for respondents and nonrespondents. If the differences are large or if the survey response rates are low, adjustments are used in an attempt to reduce the bias due to nonresponse.

A weighting-class method was used to adjust the NLS student weights for questionnaire nonresponse, but not for item nonresponse within completed questionnaires. Different response rates for students in different weighting classes were reflected in the adjustments. The method involved partitioning the entire sample (respondents and nonrespondents) into weighting classes—homogeneous groups of students with respect to the survey classification variables.

1. Classifier Variables

The five classifier variables used to define weighting classes were:

Race—White or non-White;
Sex—male or female;
High school curriculum—general, academic, or vocational/technical;
High school grades—B or better, or C or below; and
Parents' education—less than high school graduate, high school graduate, some beyond high school, or college graduate.

For the values of the classifier variables, the SRIF's, Base-Year Student Questionnaires, and First Followup Questionnaires were potential sources. (Sex was determined from the student name if no other information was available.) Values for all 5 variables were determined for over 84 percent of the 23,451; values for at least 3 variables were determined for 95 percent. The most frequent value was imputed (for weighting) for missing values.

The crossing of the 5 classifier variables yielded a table with 96 cells (2 x 2 x 3 x 2 x 4). The combining of cells with fewer than 20 respondents with similar cells yielded 87 weighting classes. Within each weighting class, the respondents' unadjusted weights were divided by the weighted response rate for that class to get the nonresponse adjustments. Six sets of adjusted weights were calculated—each set appropriate for analyses using a particular data set described in section 3 below.

2. Weight Adjustment Formulae

For each student-j selected from school-hi and classified in weighting class-k (k = 1,2, ..., 87), the K_{hij} and L_{hij} were defined:

K_{hij} = 1 if student-hij belongs to weighting class-k, or 0 otherwise; and

L_{hij} = 1 if student-hij completed the questionnaire(s) to be in data set-ℓ, or 0 otherwise.

Then the nonresponse weights for data set-ℓ (ℓ = 1,2, ...,6) were computed as

$$W_{hij(k)\ell} = W_{uhij}\, L_{hij} \frac{\sum_{hij} K_{hij} W_{uhij}}{\sum_{hij} K_{hij} L_{hij} W_{uhij}} .$$

3. Adjusted Weight Sets

Table 6-1 shows the number of sample students in each of 5 response groups, determined by which questionnaires the students completed: 16,683 (groups I and II) completed the Base-Year Student Questionnaire; limited base-year data were collected using Form B during the first followup survey from additional sample students (resurvey respondents) who had not completed the Base-Year Student Questionnaire. Thus, data for base-year items 2, 5, 8, 10, 16, 27, 83, 84, 88, 91, 92, 93, 94, and 95 (items 86-99, First Followup Questionnaire Form B) were obtained from 21,222 students (groups I, II, and IV). Two sets of nonresponse adjusted student weights were computed for analyses of base-year data—W2 for analyses of the 14 base-year items listed above and W1 for all other base-year questionnaire items.

There were 21,350 respondents (groups I, III, and IV) to the First Followup Questionnaire. The W3 weights are appropriate for analyses of items from this questionnaire only.

Analyses of change (or transition) variables derived using both base-year and first followup items should be carried out using either the W4 or the W5 weights: W5 for the change variables defined using base-year items 2, 5, 8, 10, 16, 27, 83, 84, 88, 91, 92, 93, 94, and 95; W4 for those based on other base-year items in conjunction with first followup items. Matched base-year and first followup data are available for 15,635 students (group I) for most variables and for 20,174 students (groups I and IV) for the 14 base-year variables listed above.

Table 6-1.--Groups defined by questionnaire responses

Response group	Completed Base-Year Student Questionnaire	Completed First Followup Form A	Completed First Followup Form B	Number of students
I	yes	yes	no	15,635
II	yes	no	no	1,048
III	no	yes	no	1,176
IV	no	no	yes	4,539
V	no	no	no	1,053

The last set of weights (W6) is appropriate for analyses involving every student on the NLS release tape. There are 22,398 student records (groups I, II, III, and IV) on the tape—one for each student who completed a Base-Year Student Questionnaire and/or a First Followup Questionnaire. The remaining 1,053 students (group V) who did not complete either questionnaire (but are in the sample) do not appear on the release tape. However, they do appear on the master data file.

The user should choose the set of adjusted weights which is appropriate to the data set and variables to be analyzed. Table 6-2 lists the data set and variables appropriate to the weight sets—W1, W2, W3, W4, W5, and W6. For each weight set, the adjusted weights for nonrespondents are 0 and the sum of the respondents' adjusted weights equals the sum of the unadjusted weights for the entire sample.

As mentioned previously, the weights are adjusted only for questionnaire nonresponse, not for item nonresponse. The same method could be used to obtain another set of weights, adjusted for both questionnaire and item nonresponse, for any questionnaire item or variable. In most cases, the gains from such an adjustment would not be expected to outweigh the additional cost.

Table 6-2.--Weights appropriate for data sets and variables

Weight set	Response-group data set	Survey variables
W1	Respondents to Base-Year Student Questionnaire (B-Y)	Variables defined by items from the B-Y
W2	Respondents to the B-Y or the First Followup Questionnaire Form B (FFU-B)	Variables defined by B-Y items 2, 5, 8, 10, 16, 27, 83, 84, 88, 91, 92, 93, 94, and 95
W3	Respondents to FFU	Variables defined by FFU items
W4	Respondents to the B-Y and the FFU-A, or the FFU-B	Change variables defined by items from both the B-Y and the FFU
W5	Respondents to either the B-Y and the FFU-A, or the FFU-B	Change variables defined by B-Y items 2, 5, 8, 10 10, 16, 27, 83, 84, 88, 91, 92, 93, 94, and 95 *and* the FFU items
W6	Respondents to either the B-Y or the FFU	Variables with values defined from data available for each student in the release file

☆U.S. GOVERNMENT PRINTING OFFICE: 1977—241-055:2039

Appendix

BIBLIOGRAPHY OF EXTERNAL SOURCES

Alexander, Karl, and Bruce K. Eckland. 1973. *Effects of Education on the Social Mobility of High School Sophomores Fifteen Years Later (1955-70).* Final report on project No. 10202, National Institute of Education Grant OEG-4-71-0037.

Blau, P. M., and O. D. Duncan. 1967. *The American Occupational Structure.* New York: John Wiley and Sons.

Coleman, J. S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfield, and Robert L. York. 1966. *Equality of Educational Opportunity.* Department of Health, Education, and Welfare, Office of Education. Washington, D.C.: Government Printing Office.

Duncan, O. D. 1961. "A Socioeconomic Index for All Occupations," in *Occupations and Social Status* by A. J. Reiss, Jr. New York: Free Press of Glencoe, pages 109-138.

Educational Testing Service. 1973. *The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972-Final Report.* Princeton, New Jersey.

Featherman, D. L., Michael Sobel, and David Dickens. 1970. "A Manual for Coding Occupations and Industries into Detailed 1970 Categories and a Listing of 1970 Basic Duncan Socioeconomic and NORC Prestige Scores," Working Paper 75-1. Center for Demography and Ecology, University of Wisconsin-Madison.

Flanagan, John C., and William W. Cooley. 1966. *Project Talent One-Year Follow-Up Studies.* University of Pittsburgh.

Horvitz, D. G., and D. J. Thompson. 1952. "A Generalization of Sampling Without Replacement From a Finite Universe." *Journal of American Statistical Association*, Vol. 47, pages 663-85.

Kendall, M., and A. Stuart. 1963. *Advanced Theory of Statistics.* Vol. 1. London: Charles Griffin and Company Limited.

Kish, Leslie. 1965. *Survey Sampling.* New York: John Wiley and Sons.

Seeger, Paul. 1970. "A Method of Estimating Variance Components in Unbalanced Designs." *Technometrics 12*, pages 207-218.

Sewell, W. H., and R. M. Hauser. 1975. *Education, Occupation, and Earnings.* New York: Academic Press.

Super, D. E., R. S. Kowaski, and E. M. Gotkin. 1967. *Floundering and Trial After High School.* Career Pattern Study Monograph IV. New York: Columbia University.

Trent, James W., Claire Rose, Ann Salyard, Judd Adams, Alfred C. Marcus, J. Ward Keesling, and Arthur Gerst. 1972. *An Analytical Review of Longitudinal and Related Studies as They Apply to the Educational Process.* Los Angeles: Center for the Study of Evaluation, UCLA.

WESTAT, Inc. 1972. *Sample Design for the Selection of a Sample of Schools with Twelfth-Graders for a Longitudinal Study.* Rockville, Maryland.