DOCUMENT RESUME

ED 304 988 HE 022 292

AUTHOR Knight, Shirley; And Others
TITLE Sophomore Cohort Postsecondary Education Transcript Study. High School and Beyond. Technical Report.
INSTITUTION National Opinion Research Center, Chicago, Ill.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO CS-88-675
PUB DATE Jul 88
CONTRACT 300-84-019
NOTE 83p.; Data Series: SP-HSB-87-1.0.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Academic Records; College Bound Students; *Computer Uses in Education; Data Analysis; Databases; Data Collection; *Data Processing; Higher Education; High School Graduates; Recordkeeping; Research Reports; *Student Records; Surveys
IDENTIFIERS *High School and Beyond (NCES); Postsecondary Education Transcript Study

ABSTRACT

Procedures are described that were used to collect and process postsecondary school transcripts for a subsample of members of the younger (i.e. 1980 sophomore) cohort of the High School and Beyond study who attended postsecondary institutions at any time after leaving high school. Five chapters are as follows: (1) introduction (the National Center for Education Statistics Longitudinal Studies Program, relationships between High School and Beyond and the National Longitudinal Study of the High School class of 1972, history of High School and Beyond, related studies, other files, and scope of the Postsecondary Education Transcripts Study); (2) data collection (objectives, mailings, results); (3) data preparation (objectives, data organization, computer-assisted data entry (CADE), and data quality management); (4) data processing (machine editing, organization and content of the data file, merging records, and a cautionary note on the use of credits and grades data in the postsecondary transcripts database); and (5) sample design and implementation (base year sample design, 1980 sophomore cohorts sample design, the Senior and the Sophomore Cohort Postsecondary Education Transcript Study sample, sample weights, and standard errors and design effects). Exhibits include research design for National Education Longitudinal Studies and sample CADE screens. Appended are (1) a list of endorsing institutions with the content of school transcript request packages, and (2) a list of course subject codes in numerical order. Tables are included.(SM)

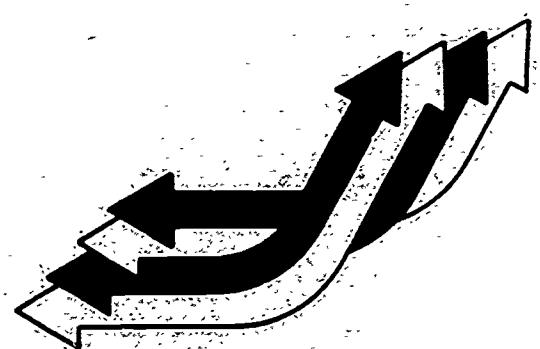
HE

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report **July 1988**

High School and Beyond
Sophomore Cohort Postsecondary Education Transcript Study

Contractor Report



Data Series:
SP-HSB-87-1.0

U.S. Department of Education
Office of Educational Research and Improvement **CS 88-675**

Prepared for the Department of Education, National Center for Education Statistics contract 300-84-019 with the U.S. Department of Education. Contractors undertaking such projects are encouraged to express freely their professional judgment. This report, therefore does not necessarily represent positions or policies of the Government, and no official endorsement should be inferred. This is released as received from the contractor.

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report **July 1988**

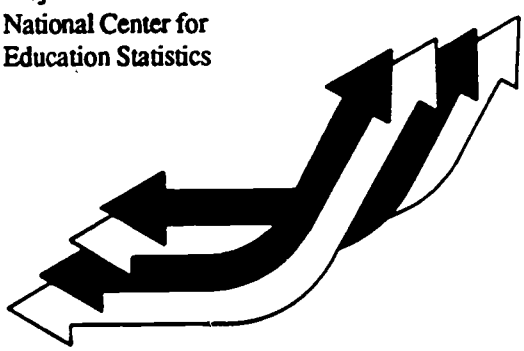
High School and Beyond

Sophomore Cohort Postsecondary Education Transcript Study

NORC
University of Chicago
1155 East 60th Street
Chicago, Illinois 60637

Shirley Knight
Penny Sebring
Martin Glusberg
Edwin Hunt
Bruce Spencer

Paula Knepper
Project Officer
National Center for
Education Statistics



Data Series:
SP-HSB-87-1.0

U.S. Department of Education
Office of Educational Research and Improvement **CS 88-675**

Acknowledgments

The authors wish to thank all those persons who contributed to the production of this report. For their considerable effort we would like to thank Tucker Landy, Hyman Bern, and Gloria Rauens in NORC's Center for Computing and Information Systems. Special recognition is extended to Shirley Walker for her assistance in coordinating the production activities. Finally, we would also like to thank those members of the NCES staff who have worked closely with us on this project: Paula Knepper, Project Officer, and C. Dennis Carroll, Chief of the Longitudinal Studies Branch.

PREFACE

The purpose of this technical report is to document the procedures used to collect and process postsecondary school transcripts for a subsample of members of the younger (i.e., 1980 sophomore) cohort of High School and Beyond who attended postsecondary institutions at any time after leaving high school. The following outline provides a general guide to the contents of the report.

- Chapter 1 contains an introduction to the longitudinal studies program administered by the National Center for Education Statistics of the U.S. Department of Education; it also describes the scope of the transcript study.
- Chapter 2 summarizes the procedures used to collect transcript data from educational institutions.
- Chapter 3 describes the Computer-Assisted Data Entry (CADE) program with which transcripts were coded and converted to machine-readable form.
- Chapter 4 includes a discussion of data editing procedures.
- Chapter 5 describes the procedures used to construct sampling weights for use in computing population estimates.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

EXHIBITS:

TABLES:

APPENDICES:

1. INTRODUCTION

The High School and Beyond (HS&B) Postsecondary Education Transcript Study, conducted in 1987, involved the collection and processing of school transcripts for a subsample of the members of the HS&B younger cohort--that is, the study's 1980 sophomores--who had attended any form of postsecondary institution since leaving high school. Transcripts were requested from schools reported by sample members in their responses to the HS&B second follow-up (1984) and third follow-up (1986) surveys. Records were obtained from all types of postsecondary institutions, ranging from those offering short-term vocational or occupational programs through major universities with graduate programs and professional schools. Information from the transcripts, including terms of attendance, fields of study, specific courses taken, and grades and credits earned, was coded and processed into a system of data files designed to be merged with HS&B questionnaire data files.

The purpose of the Postsecondary Education Transcripts Study is to provide reliable and objective information about the types and patterns of postsecondary courses taken by HS&B sample members since the base year data were collected in 1980. Because the transcript data file supplements a large, expanding database from the HS&B survey, course-taking patterns and performance can be statistically related to a wide range of other factors, including student characteristics and occupational and economic outcomes.

1.1 Overview

1.1.1 The NCES Longitudinal Studies Program

The mandate of the Department of Education National Center for Education Statistics (NCES), formerly the Center for Education Statistics (CES), includes the responsibility to "collect and disseminate statistics and other data related to education in the United States" and to "conduct and publish reports on specific analyses of the meaning and significance of such statistics" (Education Amendments of 1974-Public Law 93-380, Title V, Section 501, amending Part A of the General Education Provisions Act).

Consistent with this mandate and in response to the need for policy-relevant, time-series data on nationally representative samples of high school students, NCES instituted the National Education Longitudinal Studies (NELS) program, a continuing long-term project. The general aim of the NELS program is to study longitudinally the educational, vocational, and personal development of high school students, and the personal, familial, social, institutional, and cultural factors that may affect that development.

The overall NELS program uses longitudinal, time-series data in two ways: (1) each cohort was surveyed at regular intervals over a span of years, and (2) comparable data were obtained from successive cohorts, permitting studies of trends relevant to educational and career development and societal roles. Thus far, the NELS program consists of two major studies: The National Longitudinal Study of the High School Class of 1972 (NLS-72) and High School and Beyond (HS&B). (A third major study, the National Education Longitudinal Study of 1988,

known as NELS:88, began in 1988 and will continue throughout the decade of the 1990s.)

The first major study, NLS-72, began with the collection of comprehensive base year survey data from approximately 19,000 high school seniors in the spring of 1972. The NLS-72 first follow-up survey added to the sample nearly 4,500 individuals who had been unable to participate at the time of the base year survey. Three more follow-up surveys were conducted in the fall and winter of 1974, 1976, and 1979, using a combination of mail surveys and personal and telephone interviews. The fifth follow-up survey was fielded during the spring of 1986.

The second major survey, HS&B, was designed to inform federal and state policy in the decade of the 1980s. HS&B began in the spring of 1980 with the collection of base year questionnaire and test data on over 58,000 high school seniors and sophomores. The first follow-up survey was conducted in the spring of 1982, and the second follow-up survey in the spring of 1984. The HS&B third follow-up survey was conducted in the spring of 1986.

Three survey cohorts--NLS-72 seniors, HS&B seniors, and HS&B sophomores--are displayed in Exhibit 1-1 according to their initial and subsequent survey years and their modal age at the time of each survey. As shown, the NLS-72 seniors were first surveyed in 1972 at age 18 and have been resurveyed five times since, with the last survey occurring in 1986 when these young adults were about 32 years of age. The HS&B cohorts have been surveyed at points in time that would permit as much comparison as possible with the time points selected for NLS-72. In particular, three types of comparison are possible.

First, the three cohorts may be compared on a time-lag basis (intercohort or intergenerational). For example, the high school seniors of 1972 and the high school seniors of 1980 and 1982 may be contrasted to determine changes over time in the composition, distribution, and needs of high school seniors.

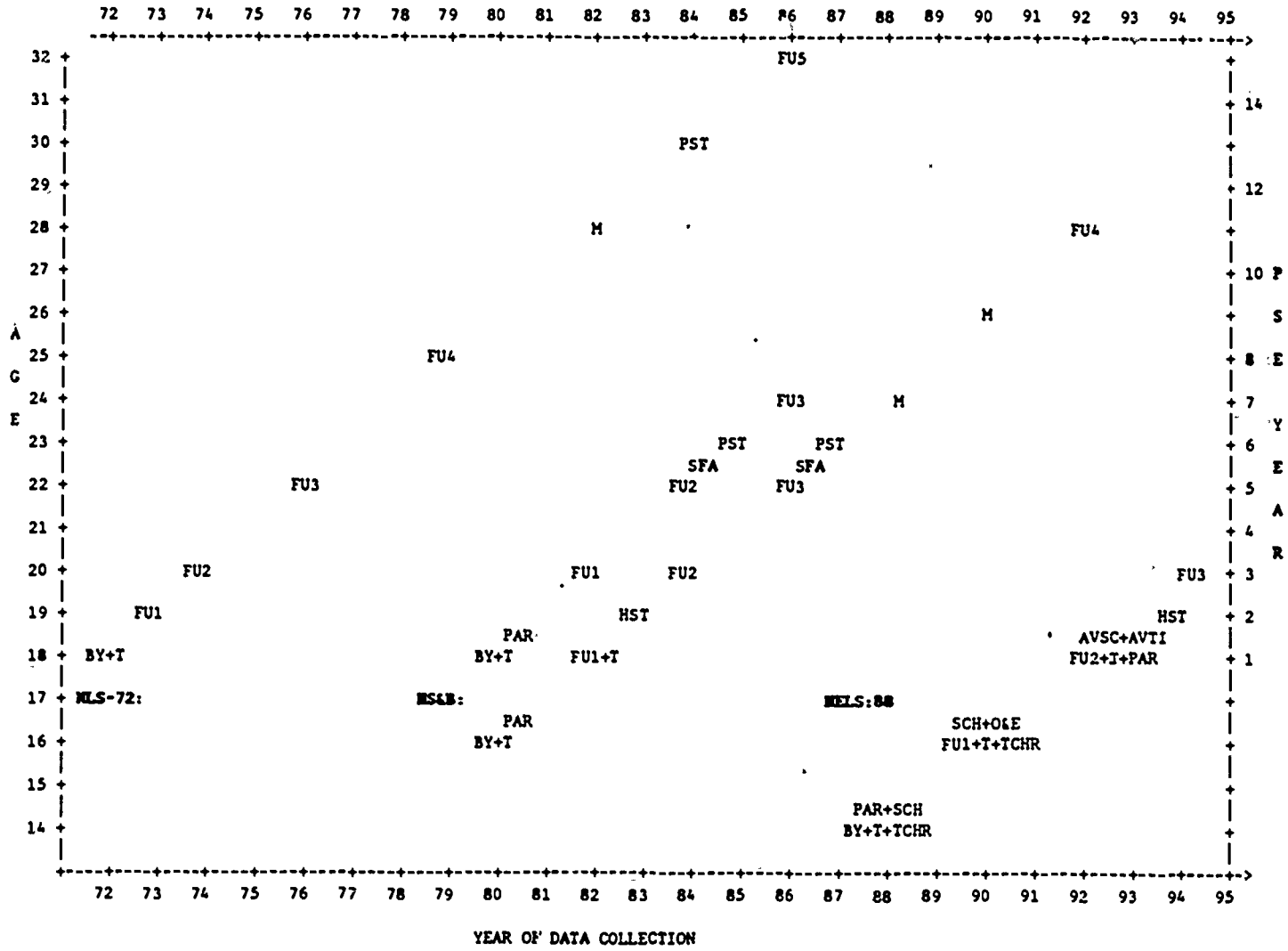
Second, fixed-time comparisons can be undertaken. For a given year, the data collection for each cohort can be viewed as a cross-sectional study. It is possible, for example, to compare employment rates in 1980 of 16-, 18- and 26-year-olds.

The third type of analysis is longitudinal (within cohort) and is designated in Exhibit 1-1 by the diagonal lines. Because the history of the age cohort can be taken into account and modeled, analyses can be designed that isolate school and program effects from the effects of differential life experiences.

1.1.2 Relationships Between High School and Beyond and NLS-72

High School and Beyond was designed to build on the NLS-72 in three ways. First, the base year survey of HS&B included a 1980 cohort of high school seniors that was directly comparable with the 1972 cohort. Replication of selected 1972 student questionnaire items and test items made it possible to analyze changes that occurred after 1972 and their relationship to new federal policies and programs in education. Second, the introduction of a sophomore cohort provided

Exhibit 1-1 Research Design For National Education Longitudinal Studies



NLS-72 = National Longitudinal Study of the High School Class of 1972
BY = Base Year Data Collection
T = Cognitive Test Administration
FU1 = First Follow-Up Data Collection
FU2 = Second Follow-Up Data Collection
FU3 = Third Follow-Up Data Collection
HS&B = High School and Beyond: 1980

TCHR = Survey of Teachers
AVSC = Area Voc. School Augmentation
AVTI = Area Voc./Technical Institute Teachers

FU4 = Fourth Follow-Up Data Collection
FU5 = Fifth Follow-Up Data Collection
M = Maintenance of Address Data
PST = Postsecondary Education Transcripts
PAR = Survey of Parents
HST = High School Transcripts
SFA = Student Financial Aid Records
NELS:88 = National Education Longitudinal Study: 1988
O&E = Offerings and Enrollments Data
SCH = School Survey

data on the many critical educational and vocational choices made between the sophomore and senior years in high school, permitting a fuller understanding of the secondary school experience and its impact on students. Finally, HS&B expanded the NLS-72 focus by collecting data on a range of life cycle factors, such as family formation behavior, intellectual development, and social participation.

1.2 History of High School and Beyond

1.2.1 The Base Year Survey

The base year survey was conducted in spring 1980. The study design provided for a highly stratified national probability sample of over 1,100 secondary schools as the first-stage units of selection. In the second stage, 36 seniors and 36 sophomores were selected per school (in schools with fewer than 36 in either of these groups, all eligible students were included). Special efforts were made to identify those students within the sample who were twins or triplets so that their co-twins or co-triplets could also be invited to participate in the study. (Data from non-sampled twins and triplets are not included in the student data files, but are available in a separate Twin Data File that links questionnaire data for both sampled and non-sampled twins for special analyses.) Over 30,000 sophomores and 28,000 seniors enrolled in 1,015 public and private high schools across the country participated in the base year survey. (Detailed information about the samples can be found in the HS&B sample design report for the base year: Martin R. Frankel, Luane Kohnke, David Buonanno, and Roger Tourangeau, Sample Design Report, NORC, 1981.)

Certain types of schools were oversampled to make the sample more useful for policy analysis. These included:

- public schools with high percentages of Hispanic students, to ensure sufficient numbers of Cuban, Puerto Rican, and Mexican students for separate analysis
- Catholic schools with high percentages of minority group students
- alternative public schools
- private schools with high-achieving students

The Hispanic supplement to the sample was funded jointly by the Office of Bilingual Education and Minority Language Affairs (OBEMLA), and the Office for Civil Rights (OCR) within the Department of Education. An additional supplementary sample was drawn from students attending Department of Defense Dependents Schools (DoDDS) located overseas. DoDDS students are not included in the data tapes distributed by NCES, however.

Survey instruments in the base year included:

- senior questionnaire
- sophomore questionnaire

- student identification pages
- a series of cognitive tests for each cohort
- school questionnaire
- teacher comment checklist
- parent questionnaire (mailed to a sample of parents from both cohorts)

The student questionnaires focused on individual and family background, high school experiences, work experiences, and plans for the future. The student identification pages included a series of items on the student's use of non-English languages, proficiency in them, and classroom experience in which those languages were used. These pages also included information that would be useful for locating the students for future follow-up surveys.

The cognitive tests measured both verbal and quantitative abilities; in addition, sophomore tests included achievement measures in science, writing, and civics, while seniors were asked to respond to tests measuring abstract and nonverbal abilities. Of the 194 test items administered to the HS&B senior cohort in the base year, 86 percent were identical to items that had been given to the NLS-72 base year respondents.

School questionnaires, which were filled out by an official in each participating school, provided information about enrollment, staff, educational programs, facilities and services, dropout rates, and special programs for handicapped and disadvantaged students. The Teacher Comment Checklist provided teacher observations on students participating in the survey. The Parent Questionnaire elicited information about how family attitudes and financial planning affected postsecondary educational goals.

1.2.2 The First Follow-Up Survey

The first follow-up sample consisted of approximately 30,000 1980 sophomores and 12,200 1980 seniors. It retained the multi-stage, stratified, and clustered design of the base year sample. All students who had been selected for inclusion in the base year survey, whether or not they actually participated, had a chance of being included in the first follow-up sample. Unequal probabilities were compensated by weighting.

A subsample of 11,500 students was selected from among the senior cohort base year participants. This subsampling was carried out to ensure adequate analytic power to address policy issues in areas such as excellence in education, access to postsecondary education, need for financial aid, and the impact of education on career choices. A special sample of 495 students was selected from among those 1980 seniors who had been selected for inclusion in the base year survey but who had not actually participated.

As in the base year survey, the Hispanic supplement to the first follow-up survey was supported by OBEMLA and OCR. In addition, the United States Army Recruiting Command (USARC) supported the retention in the first follow-up sample of 200 additional 1980 seniors who had moderate to high achievement scores but no plans for postsecondary education.

For the senior cohort, a self-administered mail-back questionnaire was the basic method of data collection. Approximately 12,200 packets containing survey questionnaires, instruction sheets, and incentive payment checks were sent to sample members during the first week of February 1982. Approximately 75 percent of the targeted senior cohort members completed and returned first follow-up questionnaires by mail. An additional 19 percent completed the questionnaires by either in-person or telephone interviews. Respondents who completed the questionnaire via telephone interview were required to have a copy of the questionnaire in front of them while doing so, to keep their survey experience as similar as possible to that of the majority of respondents, who filled out the questionnaires themselves. Follow-up interviewing was halted in mid-July of 1982, after a response rate of 94 percent had been obtained.

First follow-up data for 1980 sophomores were collected through group administrations of questionnaires and tests. The sophomore group administrations were conducted either in the sampled students' high school or in an appropriate location off campus. The location of the administration depended on the survey member's school enrollment status during the data collection period (February through May 1982). Group administrations were scheduled off-campus for sample members who were no longer attending the sampled schools. These individuals (e.g., transfer students, dropouts, early graduates) were contacted by NORC survey representatives and brought together in small groups of two to six participants. The same survey administration procedures were followed for both types of group administration. Follow-up ended in mid-July of 1982, after response rates of 81 and 89 percent had been obtained for the questionnaires and tests, respectively.

A first follow-up school questionnaire was requested of all schools selected in the base year (including those that had refused to participate), with the exception of schools that had no 1980 sophomores, that had closed, or that had merged with other schools in the sample. Schools that had received en masse transfers of students from base year schools were contacted to complete a first follow-up school questionnaire and to arrange student survey activities. These schools are not considered to be part of the probability sample of secondary schools and do not appear on the Updated School Data File. The first follow-up survey also included a sample of students from the Department of Defense Dependents Schools (DoDDS). DoDDS students were not part of the main probability sample and were not weighted.

1.2.3 The Second Follow-Up Survey

The sample design for the second follow-up survey was the same as that used for the first follow-up. Survey activities were initiated for all individuals who had participated in the first follow-up except for those who were known to be deceased.

As in the first follow-up survey, mail-back questionnaires were again the basic method of data collection for the seniors and, in this follow-up, for the sophomores as well. During the first week of February 1984, approximately 12,000 packets of survey materials were mailed to the last known addresses of the senior sample members and approximately 14,825 sophomore sample members. Extensive telephone prompting was used to encourage sample members to respond by mail. When this failed, interviews were conducted by telephone or in person.

Approximately 73 percent of the senior cohort sample members mailed back their completed questionnaires; about 13 percent were interviewed by telephone; and about 5 percent were interviewed in person. Among the sophomores approximately 73 percent mailed back their completed questionnaires; about 14 percent were interviewed by telephone; and about 5 percent were interviewed in person. As in the earlier follow-up, the survey design required that respondents who were to be interviewed over the telephone or in person have a copy of the questionnaire before them during the interview, to minimize bias due to the method of administration. Follow-up interviewing continued through July 1984, and resulted in a completion rate of over 91 percent for the seniors and 92 percent for the sophomores.

1.2.4 The Third Follow-up Survey

As in the second follow-up, mail-back questionnaires were the basic method of administration, supplemented by telephone and in-person interviews. During the last week of February 1986, approximately 26,820 packets of survey materials were mailed to the last known addresses of the sample members (senior and sophomore). Reminder/thank you postcards were mailed to respondents after two weeks. Telephone prompting started three weeks later. When this failed to elicit a response, an effort was made to complete the case by telephone. The final attempt was made through in-person interviewing.

Follow-up interviewing continued into September, resulting in a completion rate of 91 percent among sophomores, of 88 percent among seniors, and an overall completion rate of 90 percent.

1.3 Related Studies

In addition to the core surveys described above, a number of related studies have been undertaken. Besides the transcript study described in this manual, such studies have included the collection of the high school transcripts and postsecondary financial aid data for the HS&B sophomore cohort, and the collection of postsecondary education transcripts and financial aid data for the HS&B seniors. Data files for these studies and other HS&B data, such as parent surveys, school surveys, teacher comments, etc. are described below. Users' manuals or other forms of documentation are available from NCES for all data files. These auxiliary data files greatly expand the analytic potential of the core data sets, and researchers are encouraged to become familiar with them.

1.3.1 Other Base Year Files

The Language File contains information on each student who during the base year reported some non-English language experience, either during childhood or at

the time of the survey. This file contains 11,303 records (sophomores and seniors combined), with 42 variables for each student.

The Parent File contains questionnaire responses from the parents of about 3,600 sophomores and 3,600 seniors who are on the Student File. Each record on the Parent File contains a total of 307 variables. Data on this file include parents' aspirations and plans for their children's postsecondary education.

The Twin and Sibling File contains base year responses from sampled twins and triplets; data on non-sampled twins and triplets of sample members; and data from siblings in the sample. This file (2,718 records) includes all of the variables that are on the HS&B student file, plus two additional variables (family ID and SETTYPE--type of twin or sibling).

The Sophomore Teacher Comment File contains responses from 14,103 teachers on 18,291 students from 616 schools. The Senior Teacher Comment File contains responses from 13,683 teachers on 17,056 students from 611 schools. At each grade level, teachers had the opportunity to answer questions about HS&B sampled students who had been in their classes. The typical student in the sample was rated by an average of four different teachers. These files contain approximately 76,000 teacher observations of sophomores and about 67,000 teacher observations of seniors.

The Friends' File contains identification numbers of students in the HS&B sample who were named as friends of other HS&B sampled students. Each record contains the IDs of sampled students and IDs of up to three friends. Linkages among friends can be used to investigate the sociometry of friendship structures, including reciprocity of choices among students in the sample, and to trace friendship networks.

1.3.2 Other Special Studies Files

The High School Transcript File describes the course-taking behavior of 15,941 sophomores of 1980 throughout their four years of high school. Data include a six-digit course number[1] for each course taken, along with course credit, course grade, and year taken. Other items of information, such as grade point average, days absent, and standardized test scores, are also contained on the file.

The Offerings and Enrollments File contains school information, course offerings, and enrollment data for 957 schools. Other information, such as credit offered by the school, is also contained on each record.

The Updated School File contains base year data (966 completed questionnaires) and first follow-up data (956 completed questionnaires) from 1,015 participating schools in the HS&B sample. First follow-up data were

---

[1] Corresponds with descriptions in A Classification of Secondary School Courses (CSSC), developed by Evaluation Technologies, Inc., under contract with NCES, July 1982.

requested only from those schools that were still in existence in the spring of 1982 and had members of the 1980 sophomore cohort currently enrolled. Each high school is represented by a single record that includes 230 data elements from the base year school questionnaire, if available, along with other information from sampling files (e.g., stratum codes, case weights).

The Postsecondary Education Transcript File for the HS&B seniors contains transcript data on dates of attendance, fields of study, degrees earned, and the titles, grades, and credits of every course attempted at each school attended, coded into hierarchical files with the student as the highest level of aggregation. Although no survey forms were used, detailed procedures were developed for extracting and processing information from the postsecondary school transcripts that were collected for all members of the 1980 senior cohort who reported attending any form of postsecondary schooling in the first or second follow-up surveys. (Over 7,000 individuals reported over 11,000 instances of school attendance.)

The Senior Financial Aid File contains financial aid records from postsecondary institutions which respondents reported attending, and federal records of the Guaranteed Student Loan program and of the Pell Grant program.

The Sophomore Financial Aid File contains information from federal records from the Guaranteed Student Loan program and from the Pell Grant program for all students who reported attending postsecondary education and who had participated in either of these two programs.

The HS&B HEGIS and PSVD File contains the postsecondary school codes for schools HS&B respondents reported attending in the first and second follow-ups. In addition, the file provides data on institutional characteristics, such as type of institution, highest degree offered, enrollment, admissions requirements, tuition, and so forth. This file permits analysts to link HS&B questionnaire data with institutional data for postsecondary schools attended by respondents.

1.3.3 Merged Base Year and First Follow-Up Files

The First Follow-Up Sophomore File contains responses from 29,737 students and includes both base year and first follow-up data. This file includes information on school, family, work experiences, educational and occupational aspirations, personal values, and test scores of sample participants. Students are also classified as to high school status as of 1982 (i.e., dropouts, same school, transfer, or early graduate).

The First Follow-Up Senior File contains responses from 11,995 individuals and includes both base year and first follow-up data. This file includes information from respondents concerning their high school and postsecondary experiences and their work experiences.

1.4 Scope of the Postsecondary Education Transcript Studies

Although the HS&B follow-up surveys have collected longitudinal data on postsecondary educational activities of sample members, the kinds and quantity of information collected on course-taking patterns and on grades, credits, and

credentials earned has necessarily been limited by the survey methodology, and by respondents' ability to recall the details of their educational experiences.

To overcome these weaknesses and to provide a rich resource for the future analysis of occupational and career outcomes, the Postsecondary Education Transcript Study (senior and sophomore) was designed to obtain official records from academic and vocational schools. Transcript information was abstracted and coded into machine-readable form, and can thus be merged with questionnaire data and other records data (e.g., information from student financial aid records) to support powerful quantitative analyses of the impacts of postsecondary schooling.

Data files created for the transcript study include detailed information about program enrollments, periods of study, fields of study pursued, specific courses taken, and credentials earned. In addition to providing a data resource for the analysis of educational activities and their impacts, the transcript data may be used as an objective standard against which student self-reports may be compared and evaluated, thus guiding the design of future studies.

Transcript requests for the Sophomore Cohort Postsecondary Transcript Study were made for a sample of the sophomore cohort students who reported in the follow-up survey that they had attended a postsecondary institution (see Chapter 5, Sample Design and Implementation). Requests were made for 7,429 transcripts to 2,139 schools. For some of the 6,098 sampled students, multiple requests were made.

2. DATA COLLECTION

Planning for the Sophomore Cohort Postsecondary Education Transcript Study began in the winter of 1987. Preparations for data collection included three major steps:

1. Extracting information concerning each unique instance of postsecondary school attendance by younger cohort members from HS&B follow-up survey data files, and sorting this information by institution name and identification number. This data file was used to generate the printed lists of students sent to registrars and other school administrators to request transcripts.

2. Constructing up-to-date address files for all postsecondary institutions reported by sample members, and developing letters, forms, and other materials to be sent to school administrators explaining the purposes of the study, the legal authority under which the study was being conducted, and procedures for protecting the confidentiality of research subjects.

3. Obtaining the endorsement and support of a broad spectrum of professional organizations engaged in research about and representing the interests of postsecondary institutions. Appendix A contains a list of sixteen organizations endorsing the study and encouraging its members to cooperate in data collection activities.

2.1 Data Collection Objectives

The principal objective of the study was to obtain from institutions of interest reported by a sample member the formal transcripts or other equivalent records of their educational activities (i.e., documents authenticating enrollment and attendance in postsecondary programs, indicating academic or other types of performance, and showing any formal credits and credentials earned). In addition, course catalogs and other related publications were requested from these schools to facilitate the accurate and consistent coding of information about programs or fields of study, course titles, earned credits, grades, degrees or other credentials, and academic terms or other measures of enrollment duration.

A total of 7,429 transcripts were requested from 2,139 schools for 6,098 individuals (see Chapter 5). Transfer credits coded from a second school's transcripts have been systematically flagged in the data files so that analysts seeking to cumulate credits earned may easily avoid double-counting.

A secondary objective of the transcript study was to validate reports by sample members of school enrollment in their responses to follow-up surveys. Thus, transcripts were requested from each school reported in follow-up questionnaires, even if there was evidence that the respondent might not have completed the term of study or the requirements for credit. As indicated by the results described below, in a small but significant percentage of cases, institutions reported that the respondent either never actually attended classes at the named school, or else dropped out of classes before completing enough work to justify the creation of a formal record.

2.2 Mailout of Transcript Requests to Institutions

During the week of June 15, 1987, packets of transcript survey materials were mailed to the postsecondary schools. The mailing was timed to arrive at registrars' or other administrative offices at the time of the lowest level of activity for the administrative staff. The requests were received after the first activity associated with graduation and transfer and prior to expected heavy work schedules associated with fall enrollments.

Altogether, 7,429 transcripts were initially requested from 2,139 institutions for 6,098 HS&B sample members. Each transcript request package contained the following, of which examples are provided in Appendix A:

- a list of postsecondary school organizations endorsing the transcript study
- a letter to the Registrar or Director of Admissions from the NORC Director of Education Longitudinal Studies
- a letter of endorsement from the American Association of Collegiate Registrars and Admissions Officers (AACRAO)

- a letter from the Director of the Center for Education Statistics authorizing NORC to conduct the study on behalf of the Secretary of Education
- an excerpt from the Family Educational Rights and Privacy Act (FERPA) indicating the legal authorization under which the request for records was made
- a brief description of NCES's National Education Longitudinal Studies program
- general instructions for participation in the study
- a computer-generated list of students for whom transcripts were being requested
- a label to affix to each transcript to link the correct transcript to HS&B files[2]
- a transmittal form with instructions[2]
- an invoice form for transcript reimbursement[2]
- pre-paid envelopes for transcript shipment.[2]

Telephone follow-up of non-responding schools began in July when transcripts had been received from about 45 percent of the schools. Over the course of the data collection period, 1,082 follow-up calls were made to schools. Below is a breakdown to illustrate the level of effort required to obtain transcripts from a small number of schools.

Number of Schools	Number of Calls
384	1
204	2
131	3
265	4-6
98	over 6

Frequent changes of personnel, referrals to alternative administrators or sites, and problems with the typical pace of internal mail delivery systems in some schools resulted in the need to remail a total of 551 transcript-request

[2] Copies not included in the appendix.

packets. Of these, approximately 150 schools required a second re-mailing and another handful required 3 remails.

2.3 Data Collection Results

To a great degree, the success of the transcript study hinged upon the cooperation of registrars and other administrators to whom transcript requests were sent. Although 93 percent of the schools were asked to supply fewer than 10 student transcripts, 70 was the largest number from a single school. Despite the fact that transcript requests were made with the express written consent of participating subjects and photocopies were provided to schools on request, and despite the fact that study materials fully explained the legal basis for the requests for the information, school officials had the right to decline to cooperate. Most officials supported the objectives of the study, however, and were both prompt and complete in their responses. Even so, other logistical obstacles had to be overcome. A small number of schools, all in the vocational and proprietary sector, had permanently closed, eliminating access to older records. Other schools had relocated, changed their names, or merged with other institutions, necessitating extensive tracing efforts in order to deliver requests to appropriate offices, and complicating the task of locating specific student records. In the following sections we describe the response rates at three levels--the institution, the individual transcript (instance of attendance), and the student (for whom more than one transcript may have been requested).

2.3.1 The School-Level Response Rate

Transcript requests for HS&B students were sent to a great variety of postsecondary school types, including small and large private vocational and proprietary schools as well as traditional degree-granting institutions of higher education such as 2- and 4-year colleges and universities with the full range of graduate and professional programs. Identical materials and procedures were used in the collection of transcripts from all types of schools. However, as shown in Table 2-1, proportionally more non-vocational institutions (e.g., colleges and universities) participated in the study than did their vocational counterparts (e.g., trade and technical schools). The participation rates shown in the table are the simple percentages of schools in each sector that returned at least one transcript. No attempt was made in this table to adjust either for the number of transcripts requested or for the possibility that only one transcript was requested for a student who did not actually attend the school.

In the proprietary sector, only about 63 percent of the schools returned any transcripts. The sector, however, constituted only about 16 percent of the list of schools.

Schools in the other sectors were much more likely to return one or more transcripts, as is demonstrated in Table 2-2. These other types of schools constituted approximately 84 percent of the list of schools attended, and account for nearly 93.4 percent of the transcripts requested.

Table 2-1
Response Rates to the HS&B Postsecondary
Education Transcript Study By Institution Type

	Institution type						
	Proprietary	Private technical 2-year	Public technical 2-year	Public 2-year/ jr. college	Private 4-year	Public 4-year	Total schools
Percent	62.7%	84.3%	78.9%	93.1%	91.6%	95.9%	87.0%
Number of of schools in sector (N)	(341)	(89)	(157)	(479)	(608)	(465)	(2139)

The higher response rates for the public and private non-profit schools may be attributable in large part to the typically longer period during which they have been in existence, and to the relative permanence of student files they maintain. The most common reasons reported by school personnel for being unable to return transcripts were that the records had been lost or destroyed (about 8 percent of transcripts requested from schools in the proprietary sector), or that there was no record at the school of the named student's attendance (11.8 percent of the transcripts requested from these schools). An additional 1.9 percent of the proprietary transcripts requested were unavailable because of school closures. However, 15.9 percent of the proprietary schools did not respond despite assurances that they would do so.

In most cases, schools that returned transcripts also returned other related documents (e.g., bulletins and course catalogs) to assist coding.

2.3.2 The Transcript-Level Response Rate

Table 2-2 shows data collection results at the level of the individual transcript for the total sample, and separately for each of the six types of postsecondary institution. Transcript response rates are calculated as ratios of the number of transcripts received to the number of "in-scope" transcript requests. Of the 7,429 transcripts initially requested, 396 were classified as "out-of-scope" as a result of information returned by school personnel indicating that the individuals for whom transcripts were requested never attended their schools (or did not complete enough work to generate a formal record). Given this response by school administrators, these cases (transcripts) have been treated as outside the population of events being studied rather than as "missing observations." (Duplicate transcripts received from two branches of the same school were also classified as out-of-scope. They accounted, however, for less than one percent of all transcripts and had no effect on the outcome.) The

implications of this definition of "out-of-scope" transcript requests for interpreting the transcript data are discussed below.

Of the 7,033 "in-scope" transcripts requested, a total of 6,536 (92.9 percent) were returned to NORC for processing. Response rates varied from 95.4 percent for transcripts sought from public community and junior colleges to a low of 69.1 percent from the proprietary schools. Rates were uniformly high (95.4 to 94.7, 94.4 percent) from the three large strata (public community and junior and 4-year colleges and private 4-year schools). Returns were substantially lower from the strata of technical and proprietary schools.

Table 2-3 below, however, illustrates the exceptionally high rate of response at the transcript level among those schools that returned at least one transcript. The number of transcripts as a percent of those requested ranged from a low of approximately 95 percent for public technical 2-year schools to over 99 percent for the private technical 2-year schools.

As can be seen in table 2-2, reasons for non-return of transcripts varied among institution types. School refusal accounted for just under 1 percent of missing transcripts. Confirmed school closings affected only 12 transcripts. Overall, just under 2 percent of transcripts were not available because records had been lost or destroyed, but among proprietary schools 7.9 percent were in this category. The remaining category (No Response) includes transcripts from one school for which no current mailing address could be found (and which may have been closed), schools that could not be successfully contacted by telephone, and schools that expressed the intention to return transcripts but did not do so in time for processing. Also included in this category are unreturned transcripts from schools that did return a portion of the transcripts requested. Reasons for partial returns varied from clerical oversight in schools that were asked to provide large numbers of transcripts, to cases in which schools would not release a record because the student had not paid all his outstanding fees, and the like.

Table 2-2 above also shows that in 396 instances (just over 5 percent of the total of 7,429 requests), school officials reported explicitly either that the specified student had never attended the school or that the student had not stayed long enough to earn any grades or credits, and therefore had no formal records. The percentage of this type of outcome varies little across the three major strata of non-vocational or technical schools, but increases to about 14 percent of the public technical 2-year schools, and accounts for about 11 percent within the proprietary sector. For purposes of the transcript study, these cases were considered out-of-scope: they are "non-events," or at the very least they are outside of the population of events under study.

Since the initial list of instances of school attendance was created using survey responses to the HS&B second and third follow-up surveys, these results create inconsistencies between the questionnaire data files and the postsecondary transcript study data file. The discrepancy between student-reported postsecondary attendance and the evidence in school records is substantial, and so the decision to consider these instances as out-of-scope was not taken lightly. It is important to note that this status code was only assigned to cases in the survey monitoring system when school officials confirmed in writing

Table 2-2
Transcript Dispositions

	Institution Type						
	Proprietary	Private technical 2-year	Public technical 2-year	Public community/ jr. college	Private 4-year	Public 4-year	TOTAL
Received	69.1% (295)	92.1% (139)	88.7% (250)	95.4% (1,336)	94.4% (1,527)	94.7% (2,989)	92.9% (6,536)
School refused	1.6% (7)	2.0% (3)	2.8% (8)	0.6% (9)	1.0% (16)	0.7% (24)	1.0% (67)
Lost or destroyed	8.0% (34)	1.3% (2)	5.0% (14)	0.8% (11)	1.2% (20)	1.5% (46)	1.8% (127)
School closed	1.9% (8)	0.7% (1)	0.4% (1)	0.0% (0)	0.0% (0)	0.1% (2)	0.2% (12)
No response	19.4% (83)	4.0% (6)	3.2% (9)	3.1% (44)	3.4% (55)	3.0% (94)	4.1% (291)
In-scope	100.0% (427)	100.0% (151)	100.0% (282)	100.0% (1,400)	100.0% (1,618)	100.0% (3,155)	100.0% (7,033)
Never attended	11.8% (57)	7.4% (12)	14.3% (47)	4.2% (62)	5.5% (94)	3.8% (124)	5.3% (396)

their conclusion that the named student did not attend their school. Administrators had considerable information about each student named on a transcript request form, including full names, alternative names such as maiden names, social security numbers, dates of birth, and approximate dates of enrollment. In addition, there was considerable evidence in the materials returned to NORC that school personnel had conducted thorough searches for records, and often had cross-checked their results with admissions offices and financial aid offices. We therefore believe that there is little or no classification error in this status code.

One interpretation of this outcome is that HS&B respondents over reported instances of postsecondary school attendance by over 5 percent of the events (unweighted). If so, researchers analyzing postsecondary schooling using only the survey data tapes would overestimate significantly the extent of this

Table 2-3
Return Rates for Participating Schools

Institution Type	Transcripts Requested	Transcripts Received	Percent
Proprietary	302	295	97.68%
Private tech./2-year	140	139	99.29%
Public tech./2-year	262	250	95.42%
Public 2-yr./Jr. coll.	1,355	1,336	98.60%
Private 4-year	1,561	1,527	97.82%
Public 4-year	3,056	2,989	97.81%
TOTAL	6,676	6,536	97.9%

activity. Furthermore, the true discrepancy may be even bigger than that estimated by these results. In approximately half of the 293 cases in the "No Response" category of Table 2-2, neither transcripts nor any other information about the students' status was returned. In the absence of specific information to the contrary, these cases have been treated as probable instances of attendance, and therefore within the scope of the population of interest. It isreasonable to expect that if information had been obtained for these cases, some portion would have been declared as errors in reported attendance.

The fact that the rate of "Never Attended" classifications is twice as high among proprietary and public technical/2-year schools as in other sectors is consistent with descriptions of the incidence of last-minute withdrawals and dropout rates at these schools, adding face validity to this view.

However, we do not believe that the evidence is strong enough to rule out alternative interpretations. One reasonable possibility is that many of these instances of reported attendance result from errors in the coding of incomplete or marginally legible school names written by respondents into survey questionnaires. Conceivably, then, respondents may have in fact attended some form of postsecondary school, but the data in the questionnaire files may be wholly or partially inaccurate for these individuals. If this were true for each discrepant case, then the questionnaire files would accurately reflect the extent of postsecondary educational activity, but would include measurement errors concerning the specific school attended.

A third alternative seems to us equally persuasive. Although there were 396 transcript classifications of non-attendance, only 229 individual sample members were classified as out-of-scope as a result. Of these 396 transcripts, 58.3

percent (231) were requested for the 229 out of-scope students.[3] However, the 229 students represent only 3.8 percent of the total sample of individuals (6,098). Only one transcript was requested for 227 of these individuals and two transcripts each for two individuals. For these 229, school officials returned a report of no or insufficient attendance.

Although a detailed analysis has not been possible, it is conceivable that many of these individuals may have attempted to report the same institution in both HS&B follow-up surveys, but in one instance returned a low-quality response resulting in a coding error for one of the two reports. If a transcript was returned for the correctly coded school, a thorough analysis of its contents could shed considerable light on the nature of the apparently erroneous report. The contract for data collection and processing did not include support for this type of analysis. However, the public use data files contain data records for all 6,098 sample members for whom transcripts were requested (including the 229 classified as out-of-scope), and include separate records for each transcript requested (including the school identifiers for the 396 transcripts classified as out-of-scope), thus providing researchers with all the material needed to fully assess the issues of measurement error. The variable FINDISP stored on each of the transcript-level records identifies out-of-scope transcripts (and sample members) for further analyses.

Researchers should note, however, that the adjusted weights attached to the transcript file apply to individuals, not transcripts. Thus, adjusted weights are attached only to the 5,869 "in-scope" sample members.

2.3.3 Student-Level Data Collection Results

Transcripts were sought for 6,098 selected (see Chapter 5) HS&B 1980 sophomore members who reported attending postsecondary schools since leaving high school. Reports of postsecondary attendance were obtained from HS&B second and third follow-up survey questionnaire responses. To be eligible for the transcript study, respondents must have provided specific information (i.e., the name and, desirably, the city and state) about at least one of the postsecondary schools attended. As described above, reports from school personnel indicated that 229 individuals who reported attending only one or two postsecondary schools had not actually attended those schools (or had not completed enough work to have established a formal record).

Table 2-4 presents distributions of the number of transcripts received for each student. Excluding the 229 out-of-scope cases, one or more transcripts were obtained for 94.3 percent of the 5,869 enrollees. A single transcript was received for 4,620 cases (78.7 percent of this group). Two transcripts were processed for 829 individuals (14.1 percent) and three or more transcripts were obtained for 84 sample members (1.5 percent).

---

[3] Multiple transcripts were requested for many individuals for whom some or all transcripts may have out-of-scope. Thus, an individual could have both in-scope and out-of-scope transcripts requested for them. See Section 2.3 for further detail.

Table 2-4
Number of Transcripts Received: HS&B Postsecondary Education Transcript Study

Number of transcripts	Number of respondents	Percent of in-scope respondents	Percent of total respondents
None (in-scope)	336	5.7	5.5
One	4,620	78.7	75.8
Two	829	14.1	13.6
Three	78	1.4	1.3
Four	6	.1	.1
Total in-scope	5,869	100.0	96.3
None (out-of-scope)	229	NA	3.8
TOTAL SAMPLE	6,098	96.2	100.1

In addition to collecting multiple transcripts per case, many transcripts contained information about credits transferred from other schools. Transfer credits were specially flagged in the data files to assist researchers in avoiding double-counting of earned academic credits by those who attended more than one school. Transfer credits for 5,533 individuals have been documented in their transcript records. The variables TRNSFERS on the student-level record and TRNSFERT on the transcript-level record in the data files identify individuals and transcripts containing transfer credits.

3. DATA PREPARATION

3.1 Data Preparation Objectives

The diversity in structure and contents that exists among the transcript records reflects the great variability among the schools from which they were obtained. Although transcripts from public and private 2-year and 4-year colleges were generally similar with respect to the data they contained, for example, they nevertheless differed in their physical layout and in the terminology used for identical or related concepts. Early in the design stage for the Senior Cohort Postsecondary Transcript Study, it became apparent that the superficial similarities in many transcripts give way to countless differences in the ways in which academic progress is measured and recorded. This is especially true of course grades and credits.

The variability across institutions in the details of transcript information defies any simple reconstruction or homogenization. Virtually any element on an academic transcript, including such seemingly straightforward items as course titles, may be subject to highly particularized local conventions whose logic may

be independent of, or even contravene, common practices. For example, it is not uncommon to find courses in English composition merged with other content and carrying formal names suggesting that they belong in the social science curriculum. Such instances, by no means rare, were resolved by Computer-Assisted Data Entry (CADE) staff, who consulted program-of-study catalogs and descriptions of courses obtained from the postsecondary institutions.

Even more problematic was the issue of standardizing metrics for such typical transcript elements as grades or credits. For example, the notion that one school's grading or credit system can be equated to another's by a simple linear transformation of scores may have been defensible for secondary school grades in the sophomore cohort high school transcript study. Attempting the same sort of "equating" with postsecondary school grades and credits carries the risk of introducing systematic errors into complex analyses.

In preparing the data for conversion to standardized, machine-readable data files, NORC's approach was to impose a common structure and organization on the transcript information, but to preserve to the extent feasible the actual information contained in the original documents. Thus, grades and credit values are stored as they were reported, and have not been transformed to any common metric. Such fields as degrees and credentials earned, major and minor fields of study, and titles of courses taken have been assigned numeric codes as explained below, but also have been recorded exactly as they were reported on the transcripts.

This approach places some additional burden upon transcript data users to gain familiarity with the variability across institutions and sectors in the data values stored in such fields as grade point averages, course grades, and credits. Our exposure to these data during their collection and processing leads us to conclude that in order to use these complex files effectively in educational research, each analyst should make a detailed assessment of the properties of all transcript elements of interest.

As is described in detail below, data preparation was carried out by a staff of 10 specially trained coders under the guidance of a supervisor and the data preparation manager. The data preparation task included analyzing the transcript document to determine its general organization and special characteristics, abstracting standard information from the highly varied documents into a common format, and assigning standard numerical codes to such transcript data elements as major and minor fields of study, degrees earned, types of academic term, titles of courses taken, grades, and credits.

3.2 Data Organization

Transcript data were organized into a four-level hierarchy consisting of data at the student, transcript, term and course levels. (See Exhibit 3-1.) At least one student-level and one transcript-level record is provided for each sample member for whom a transcript was requested, even if the school reported that the individual had never attended, or had withdrawn before establishing a formal record. Records in this category are flagged with a special disposition code. (See Chapter 2 above for a discussion of out-of-scope cases.)

Student-level data refer to general information about the responde 's educational career. All records are assigned case ID codes, allowing merger of transcript data with other files (term and course), relevant questionnaire data from the HS&B base year and follow-up surveys (e.g., self-reported high school program), high school grades, composite and derived variables from survey data, (base year SES quantities, achievement test quantities, etc.), data on the respondent's high school, sampling weights, and data summarizing information found on transcripts for all postsecondary schools attended (e.g., an educational activity status measure for several points in time between 1981 and 1987).

Transcript-level records contain data pertaining to a student's academic record at a single institution, including the institutional ID code (FICE code or vendor number), degree(s) or other credentials conferred with accompanying dates, major and minor field(s) of study, and the student's cumulative grade point average (GPA).

The term-level of the hierarchy contains information describing specific units of instruction. Term records usually refer to commonly understood academic terms such as quarters, trimesters, or semesters. Term-level records include the the type of term, season, start and end dates, the type and characteristics of the grade scale employed during the term (e.g., letter or numeric scoring), the number of courses associated (and hence the number of course-level records attached) with a term, and a special flag indicating regular or transfer status for the term. The term type flag includes a code denoting credit for major standardized tests (e.g., CLEP, LSAT) as well as work and other life experiences for which credit is given.

Course-level records store the data for each course taken by a student during a specific term. The formal title of the course was entered verbatim from the transcript, then assigned one of 78 academic or vocational program codes based on those contained in the publication, A Classification of Instructional Programs(Malitz, G.S., et al.; Washington, D.C.: National Center for Education Statistics, U.S. Department of Education, 1981, hereinafter referred to as "CIP"). The 78 instructional program codes employed in this study included 41 major program areas (2-digit), 20 program sub-groupings (4-digit) and 17 individual programs (6-digit). An additional code was reserved to indicate lump-sum transfer course credit. A list of the 78 program classifications and their related CIP codes is included as Appendix B. Also entered were credits attempted and the grade received by the student for each course.

3.3 Computer-Assisted Data Entry (CADE)

3.3.1 CADE Concept

In a conventional survey, the major data preparation tasks, editing/coding and data conversion, are performed in sequence by different individuals. The editor-coder follows a set of defined procedures to select, classify, and systematize data. The edited and coded documents are then given to data conversion operators for efficient, accurate conversion of the data to machine-readable form. Usually, the training and skills most appropriate for a coder differ considerably from those of a data conversion operator.

EXHIBIT 3-1
NS&B Transcripts Study: Data Organization

I. Student-level record
- Student ID
- Numbers of transcripts requested
- Numbers of transcripts received
- Transcript data indicator
+ Transfer courses flag
- Survey data and composite variables from student data files:
  - Socio-demographic variables
  - Characteristics of secondary school attended
  - Base year and follow-up study test scores
+ Postsecondary school enrollment status indicators
- Sampling weights

II. Transcript-level record
* Student ID
* School ID (FICE or vendor number)
- IPEDS number
+ Final disposition of transcript requests
- Postsecondary school census region
- Postsecondary institution type
- Sequence number
+ Number of terms per transcript
* Degree awarded:
  - Type of degree
  - Verbatim degree text
  - Date degree conferred (month and year)
* Cumulative grade point average
* Field(s) of study:
  - Verbatim text-major
  - Major instructional program code
  - Verbatim text-minor
  - Minor instructional program code

III. Term-level record
+ Student ID
+ School ID (FICE or vendor number)
+ Transcript number
+ Term number within transcripts by SORTDATE
* Date of term (month or season and year)
* Institutional context of term (transfer or non-transfer term flag)
* Type of term:
  - Types of academic terms
    - Quarter, trimester, semester, variable length
  - Types other than academic terms
    - Test terms, other than test terms
* Grade scale type in effect during term:
  - Letter grade scale
  - Numeric grade scale
    - Highest grade possible
    - Lowest grade possible
    - Minimum passing grade

IV. Course-level record
+ Student ID
+ School ID (FICE or vendor number)
+ Transcript number
+ Term number
* Grade received for course
  - Letter grade for course
  - Numeric (0-100) grade for course
  - Numeric (0-4) grade for course
* Credits attempted for course
* Verbatim text of course title
* Course program code

* Denotes data recorded from transcripts using CADE.
+ Denotes data derived from transcripts but not entered directly.
- Denotes data merged from other data sources.

The HS&B Postsecondary Transcripts Study required abstracting, coding, and organizing data from over 6,500 forms that varied greatly in appearance and content. Compared to the typical survey questionnaire, the amount of data to be keyed per transcript was very small. The majority of the coding task involved the assignment of Course and Major/Minor program codes, selected from a rather complex taxonomy. Previous experience on complex data abstraction studies involving small amounts of keyed data had shown that reasonable efficiency gains could be expected by combining coding and data conversion. Guided by this experience, NORC successfully modified its proprietary Computer-Assisted Data Entry (CADE) system to accommodate the data processing of postsecondary transcripts.

For the purposes of the HS&B Postsecondary Transcripts Study, a single member of the coding staff reviewed a transcript for all relevant, in-scope data, classified those data, and entered the data into a computer file. Combining these steps ensured that transcripts would be handled as internally consistent, integrated records of an individual's educational activity. Moreover, since all transcript processing occurred at a single station, the use of CADE reduced the number of steps at which records might be lost or misrouted, or other errors introduced into the database.

3.3.2 CADE Equipment: Hardware and Software

The CADE program used in this study was prepared at NORC using the fourth-generation database language Metafile on the IBM-compatible Corona microcomputer. Each of 10 CADE operators was assigned to a microcomputer station for transcript processing. The CADE program prompted the operator, through a series of defaults, for entry of all of the data elements requiring entry (i.e., all data elements marked "*" in Exhibit 3-1). The program repeated this cycle through the transcript-, term-, and course-level until all data for a transcript had been entered. Operator access to any level of the data hierarchy for revision, editing, and the like was made possible through selection menus.

Exhibits 3-2 through 3-10 illustrate entry screens that prompted the operators for entry of data at transcript-level, term-level, and course-level.

The CADE program enforced a set of predetermined range and value limitations on each field, making it impossible for CADE operators to enter, for example, an illegitimate school ID (FICE code/vendor number), student ID, or combination of the two. The program allowed entry of only the 79 predetermined CIP codes at the transcript-level (major and minor) and course-level. Similarly, grades and credit values entered had to fall within specified ranges.

The most difficult aspect of transcript coding is classifying the fields of study and formal course titles using the CIP taxonomy. The CADE operators were issued coding manuals that included CIP category dimensions, as well as course catalogs and other resource materials relevant to transcript coding. To supplement conventional uses of the CIP manual, the

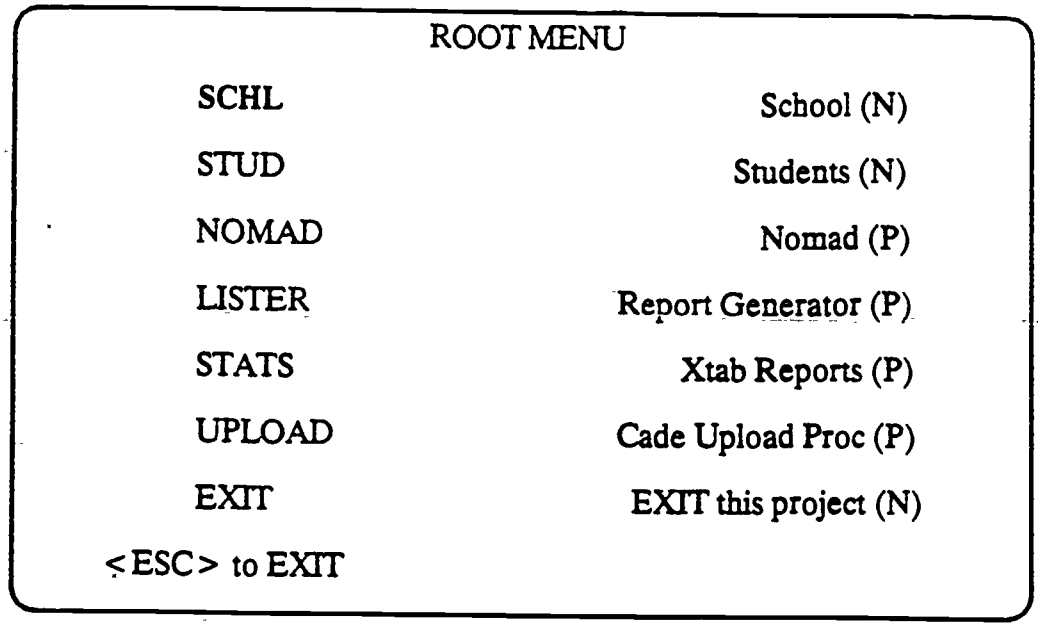


Exhibit 3-2 The initial CADE screen depicting data processing options. For entry of a transcript record, the CADE selects the "SCHL" option (highlighted).

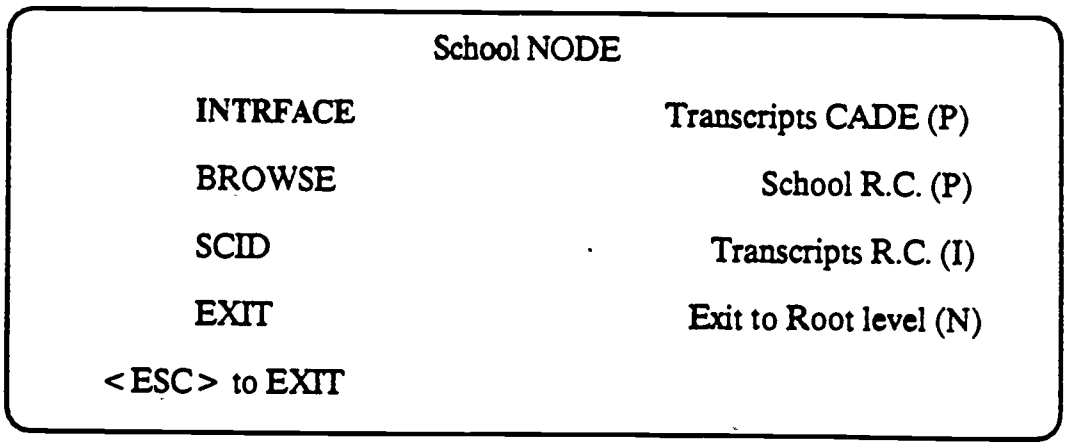


Exhibit 3-3 CADE operator selects option "INTRFACE" to begin entry of a transcript record.

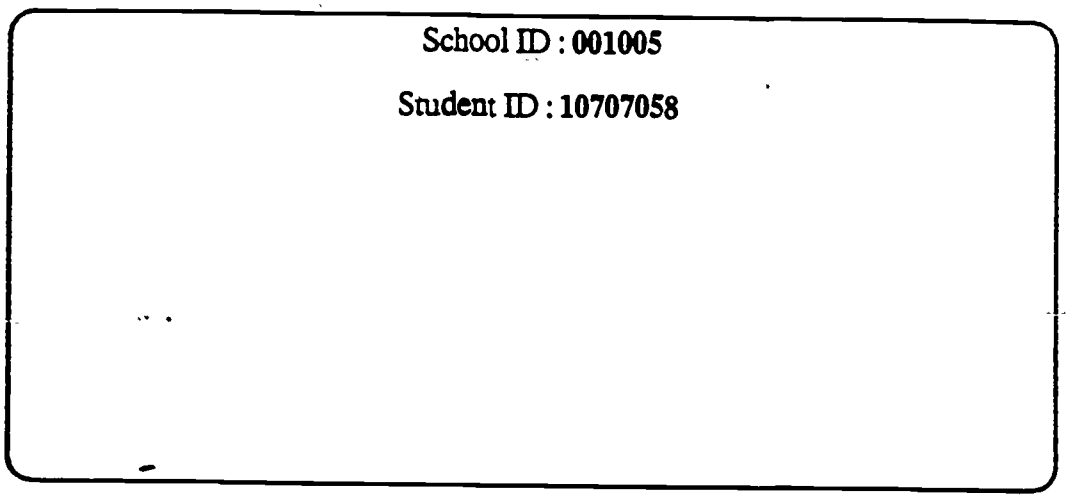


Exhibit 3-4 Operator enters valid school ID (FICE) and student ID combination.

School 001005 Student 10707058 Disp:0 Id 99 count 0 at 11:21AM

CADE main menu--select function

ENTER

EDIT

VIEW

VERIFY

QUIT

Exhibit 3-5 Main CADE options menu. The CADE operator selects the option necessary for processing data. Notice that screen includes the school, student, current disposition of record, CADE operator ID number, count (i.e., total number of terms and courses in the record), and current system time. In this case the CADE operator selects "ENTER", to enter a transcript record.

School 001005 Student 10707058 Id 99 count 0 at 11:21AM

11111111111111111111111 Enter degree information 11111111111111111111

Kind	Text	GPA	Month	Year
2	BS	4.3	12	86

Text of Major	Cip	Text of Minor	Cip
RUSSIAN HISTORY	67	COMPARATIVE LITERATURE	41

11111111111ESC = quit 111111**F1 = help**1111F2 = show CIP numbers1111111111

Kind : 1 = Assoc. 2 = Bachelor 3 = Master 4 = PhD 5 = License 6 = Cert. 7 = None
GPA : 0.0--4.0, or 9.9 = missing
Year : 81--87, or 99 = missing
Month: 1--12, or 99 = missing

Exhibit 3-6 CADE prompts the operator to enter degree-related information. Striking the F1 function key produces a listing of all valid codes for each variable at the transcript level.

School 001005 Student 10707058 Id 99 count 0 at 11:23AM

= CADE entry menu--select function =

Add Term	**(to end)**
Add Course	(to end)
Insert Course	
Quit	

Exhibit 3-7 CADE operator selects "Add Term" for entry of first term appearing on transcript.

School 001005 Student 10707058 Id 99 count 0 at 11:24AM

Enter term information

Term	Transfer	Grade Scale	Term Type	Season	Start Year
01000	1	1	2	1	81

-----ESC = quit-----**F1 = help**----------

Transfer: 1 = transfer term 2 = regular term
Scale : 1 = letter 2 = numeric 0--100 3 = numeric 0--4 8 = missing
Type : 1 = variable/non-course 2 = semester 3 = trimester 4 = quarter 5 = test 9 = miss
Season : 1 = fall 2 = winter 3 = spring 4 = summer 5 = no season 9 = unknown
Year : 81--87 or 99 = missing

Exhibit 3-8 Term data are entered into transcript record. The CADE operator summons a "help" list of all valid, term-level codes and labels by pressing the F1 key (bold).

School 001005 Student 10707058 Id 99 count 0 at 11:25AM

= CADE entry menu--select function =

Add Term	(to end)
Add Course	(to end)
Insert Course	
Quit	

Exhibit 3-9 The CADE operator now selects "Add Course", to enter the first course for Term 1.

School 001005 Student 10707058 Id 99 count 0 at 11:25AM
Enter course information

Course	Grade	Credits	Course Title	Cip
01001	A	3.0	CONTEMPORARY POETRY	41

ESC = quit edit F1 = help F2 = show CIP numbers

Term:01000 Transfer:1 Scale:1 Type:2 Season:1 Start: /81 End: /

Grade : 0 thru 100, or S,U,P,W,WP,WF,I,IP,IF,CR,AU,NO,M
Credits: 0--999.999 (999.998 = await supervisor edit/delete)
Cip : 1--78, or 95 = uncodable, or 96 = none/not applicable

Exhibit 3-10 Data corresponding to the first course in Term 01000 (1) are entered by the operator. The screen includes a view of the term-level data already entered (bold).

CADE program included a computerized version of it. When this feature was activated, coders were able to obtain a screen display of the CIP codes and their definitions.

3.3.3 CADE Operator Training

The CADE operator staff was given six days of intensive training, which included formal classroom instruction and independent coding practice and drill. Each day's training lasted a full eight hours, because of the novelty of the coding/data entry technique employed and the complexity of the task. The benefit of the training investment was immediately apparent in the high quality of the coding work (both initially and throughout the period of activity) and the exceptionally low turnover rates for the coding staff. It was also reflected in
the completion of the coding task 16 days ahead of schedule.

CADE operator training addressed the following topics:

- Hierarchical organization of transcript data
- Analysis of transcript-document formats (special emphasis on documents received from non-HEGIS institutions)
- CIP codes, dimensions of instructional program categories
- Operation of CADE using the IBM-compatible Corona PCs
- Progressive, skills-improvement drills at the PC
- Individual exercises with mock transcript coding

CADE operator trainees reviewed sample transcripts from a wide variety of school types: HEGIS and non-HEGIS, 2-year and 4-year, private and public. Drills, designed to increase coder identification of in-scope data, were conducted daily with excellent results. A major component of classroom training addressed the logic of the instructional program category dimensions and the CIP codes.

3.4 Data Quality Management

Quality control of transcript record data was introduced and maintained through a combination of procedures: error prevention features within the CADE program, verification re-entry of transcripts, supervisor analysis of course-file records, supervisor review of entire transcript records, and the continual availability of coding supervisors for consultation and guidance.

The CADE program itself screened for error in three ways. Through a check-digit system, the program disallowed entry of incorrect identification data (i.e., school FICE codes, student ID numbers, and combinations of schools and students). Furthermore, each data field was programmed to disallow entry of illogical or otherwise incorrect data. For example, a coder was automatically prevented from entering a letter grade for a course

if a numerical grading system had been specified on the term-level records under which the courses were listed.

Ten percent of each CADE operator's output was subject to verification re-entry by a trusted, specially trained verifier. The verifier was chosen by the supervisor to re-enter selected cases and note patterns of discrepancy in coding. The verification procedure enabled management to better assess the degree of agreement among coders. Verifier re-entry of transcripts involved 886 transcripts, or 13.5 percent of the transcripts processed. Of the 886 re-entered transcripts, the verifier found at least one disagreement in 565 cases, the majority of these occurring in the first three weeks.

All terms and courses were assigned to 1 of 16 course-files, to await eventual mainframe upload. A special report utility in CADE allowed management to dump all terms and courses stored in a particular course-file for critical examination. Where problems were observed, for example, in a specific category of courses, a more detailed report could be produced that showed only those courses corresponding to one or several CIP categories. Course-file analysis led to several important updates to the CADE operators' manual.

The CADE shop supervisor analyzed some 18,000 courses over a 14-week period, 10,000 of which were coded during the first and second weeks of production.

One supervisor critically reviewed 649 randomly selected transcripts (9 percent of transcripts processed). A supervisor submitted weekly reports to management detailing error rates for each variable in each hierarchy. The rate of error was calculated by dividing the number of times a given variable was coded (i.e., "chances") by the sum of errant coding decisions. The rate of error calculated for the two variables deemed most critical, major field of study CIP and course CIP, was 5.3 percent (major field of study) and 3.8 percent (course CIP).

As part of quality control, supervisors also reviewed screens of transcript records. These screens included the user-file ID of the CADE operator who entered the record, allowing the supervisor to make individual assessments and thus provide personal feedback to staff.

As unanticipated problems arose during the CADE period, a policy decisions protocol was followed. All questions and other issues were directed to project management staff for assessment and final coding decision. The resulting decisions were routinely distributed to the CADE operators, to be added to their coding manuals.

4. DATA PROCESSING

Data processing activities began with the construction of the subsample of postsecondary attenders from the main survey files, and the creation of lists of institutions from which transcripts were to be requested. They continued with the development of programs and materials to request transcripts and to monitor

data collection activities, and with the adaptation of NORC's Computer Assisted Data Entry (CADE) system for the abstraction and coding of transcript information. These activities have been described in Chapters 2 and 3 of this manual. Once transcript data was converted to machine-readable form, the data was restructured into a set of four rectangular data files for efficient storage. It was then uploaded from microcomputers to mainframe facilities, and further processing included computer editing of the data, and the creation of sets of program control files to permit the construction of analysis files using either SAS or SPSS, the two most commonly-used statistical packages for analyzing NELS data sets. Finally, two sets of adjusted sampling weights were created for making population estimates with transcript and other survey data. This chapter describes the activities from machine editing through data file construction. Sampling and weighing are the subject of Chapter 5.

4.1 Machine Editing

As described in Chapter 3 above, the CADE program was designed with extensive controls on data entry, resulting in very low error rates for all elements in the raw data. The computer editing strategy was guided by the same principles as the CADE design process--that is, a highly flexible approach was necessary to accommodate the tremendous variation in format and quality of transcripts.

To begin with, a thorough analysis was made of the distribution of values for each separate item in the raw data files. The purpose of this check was to identify data values that, based on knowledge of and experience with transcript data, appeared to be errors. Because of the extensive "front-end" cleaning performed by the CADE program, the bulk of the raw data items appeared to have very few errors, with the average error rate less than one half of one percent. In most instances, stray codes and illegal values were the results of specific keying errors that could not be prevented in a cost-effective manner by the CADE program.

4.2 Organization and Content of the Data File

The CADE program processed data at three levels described in Chapter 3: transcript, term, and course. The design of the final data files called for an additional data level, the student level, under which all transcript data for each sample member would be ordered. The student record was formed by aggregating all records for an individual student, merging data from the sampling and receipt control files, computing a series of composite variables based on data from all of a student's transcripts, and finally merging in a set of composite variables from the main HS&B third follow-up survey data tapes.

In designing the final transcripts database, data storage efficiency was a major consideration. A standard rectangular file organization was ruled out because the amount of space required to handle the maximum record length for every case would have been impractical. Further, because the amount of data stored for each case was extremely variable (most cases for 3-4 year schools had an average of 1 transcript, about 8 terms, and about 31 courses, but some cases had 3 times this amount of data), a flat file structure would have been populated with empty data fields for most cases. Vocational schools averaged 1 term.

To optimize storage space for the vast amount of information in the transcript study database, each of the four record types (student, transcript, term, and course records) was written to a separate file. Analysts may use the four files individually or jointly, depending upon their specific research objectives. For many analyses, researchers may find it sufficient to use only the composite variables from the questionnaire and transcript files stored on the student-level records. For other purposes, merging the student- and transcript-level records will provide the amount of detail desired. However, for most studies, it will be necessary to merge all four files into a single hierarchical file in which courses are nested within terms, terms within transcripts, and transcripts within students. Once this merged file is created, analysts may construct any number of composite indicators of educational activity, and then reduce the data matrix to keep only the variables essential to the analysis in a rectangular file with one record per student. Managing the data in this way will reduce storage costs for online data sets and will minimize the computing (CPU) time necessary to obtain results.

The student and transcript record data files contain information for 6,098 survey respondents in the transcript survey sample. Each member of the sample (including the "ineligible" cases described in Chapters 2 and 5) has a student-level record in the file. A transcript record was created for each requested transcript, even if the transcript was not returned, or if school officials reported that a student had never actually enrolled. Cases for whom transcripts were requested but not received have "dummy" transcript records in the file.

On each transcript record is a disposition (status) code showing either that the transcript was received and processed, and that term and course records exist in the appropriate files, or showing the reason (if known) that the transcript was not received (e.g., school had closed, records lost or unavailable). For cases (transcripts) defined as out-of-scope (see Chapters 2 and 5), this disposition field contains the code indicating that the school reported that the student never attended the named school and that no transcript exists. Researchers should note that any given sample member may have a combination of transcript records classified as "received and processed", "out-of-scope", and "not received, but in scope" associated with his or her student-level record. (For conventional analyses of these data using the adjusted weights attached to these files, it is strongly recommended that analysts first purge the files, including ineligibles and eligible nonrespondents, of those with no transcripts.)

Associated with each "received" transcript are one or more Term Records containing data for each of the terms reported. (No term or course records were created for cases for which no transcripts were received.) Separate course records were created for each unique course taken within a term, including failed and audited courses.

4.2.1 The Student Record

As noted above, a student-level record was included in the database for every sample member for whom a transcript was requested, including those who later proved to be ineligible (never attended), or for whom no transcripts were received. Student-level records contain identifying and survey control data,

activity state pointers, composite variables from the main HS&B files, and weights.

4.2.2 The Transcript Record

One transcript-level record was created for each transcript requested for each sample member. There is at least one Transcript Record for every student; over one-third of HS&B sample members have multiple Transcript Records. For ineligible cases, or for eligible cases for whom a transcript was not received, the transcript-level record is a placeholder or dummy record where information about the transcript request (e.g., the institution's ID number, the final data collection status code, etc.) is stored. If a transcript was received and processed, the transcript record stores information related to the entire period of attendance at the school, such as degree received, grade point average, whether the school accepted any transfer credits, and so forth. Information related to specific terms of attendance or specific courses taken is stored on term- or course-level records, which may be linked by a combination of ID keys to the transcript record with which they were originally associated. There are a total of 7,429 Transcript Records in the HS&B sophomore cohort database.

A total of 443 (never attended/plus duplicates) transcript-level records exist for out-of-scope cases. These records should be omitted from conventional analyses. Although raw weights have been included for these cases to permit the calculation of additional customized weights, the adjusted weights for ineligible cases is always set to the value "zero" (see Chapter 5).

4.2.3 The Term Record

A Term Record was created to store data for each term associated with a transcript, and to provide an organizing mechanism for linking course-level records associated with a given term and transcript. Students have widely varying numbers of Term Records (up to 22 terms), reflecting the amount of time spent in postsecondary schooling. Students who enrolled only in one short-term vocational program, or who stayed for only one semester at an academic institution, may have only a single Term Record in the file. Approximately 10 percent of the 6,536 coded transcripts had a single associated Term Record. Students continuously enrolled in institutions of higher education since high school graduation have many more Term Records in the database. The HS&B sophomore cohort database includes a total of 43,592 Term Records (covering the 5,533 students for whom one or more transcripts were received). Approximately half (51 percent) of the transcripts in the file are linked to four or fewer terms. An additional 32 percent of the transcripts are linked to between 5 and 8 Term Records. Eight percent of the transcripts are linked to more than 10 Term Records.

Most Term Records describe conventional academic terms of study such as semesters or quarters. These Term Records store data that pertain to courses taken during the specified term, and which otherwise would have been repetitively and wastefully stored directly on Course Records. Term Records include such items as beginning and ending dates for the term and the grade scale being used for the courses taken in that term. In some cases, grading schemes at a school

changed during a student's period of attendance. Data on term records help to identify these instances for proper handling.

4.2.4 The Course Record

One Course Record was created for every course reported on a transcript. Credit-bearing entities other than courses were also stored in course records (e.g., credits earned through work experience or by examination). Varying numbers of Course Records are associated with each term for a particular student. In all, there are 194,672 Course Records stored in the HS&B sophomore cohort database.

4.3 Merging Records

As described above, the postsecondary transcript database consists of four files, one for each record type. However, the individual record types have been designed to allow for the merging of data from two or more files into a single hierarchical file, or, if necessary, into a very large rectangular file. The relationship among the various record types and the identifiers needed to merge levels are summarized in Exhibit 4-1.

4.4 A Cautionary Note on the Use of Credits and Grades Data in the Postsecondary Transcripts Database

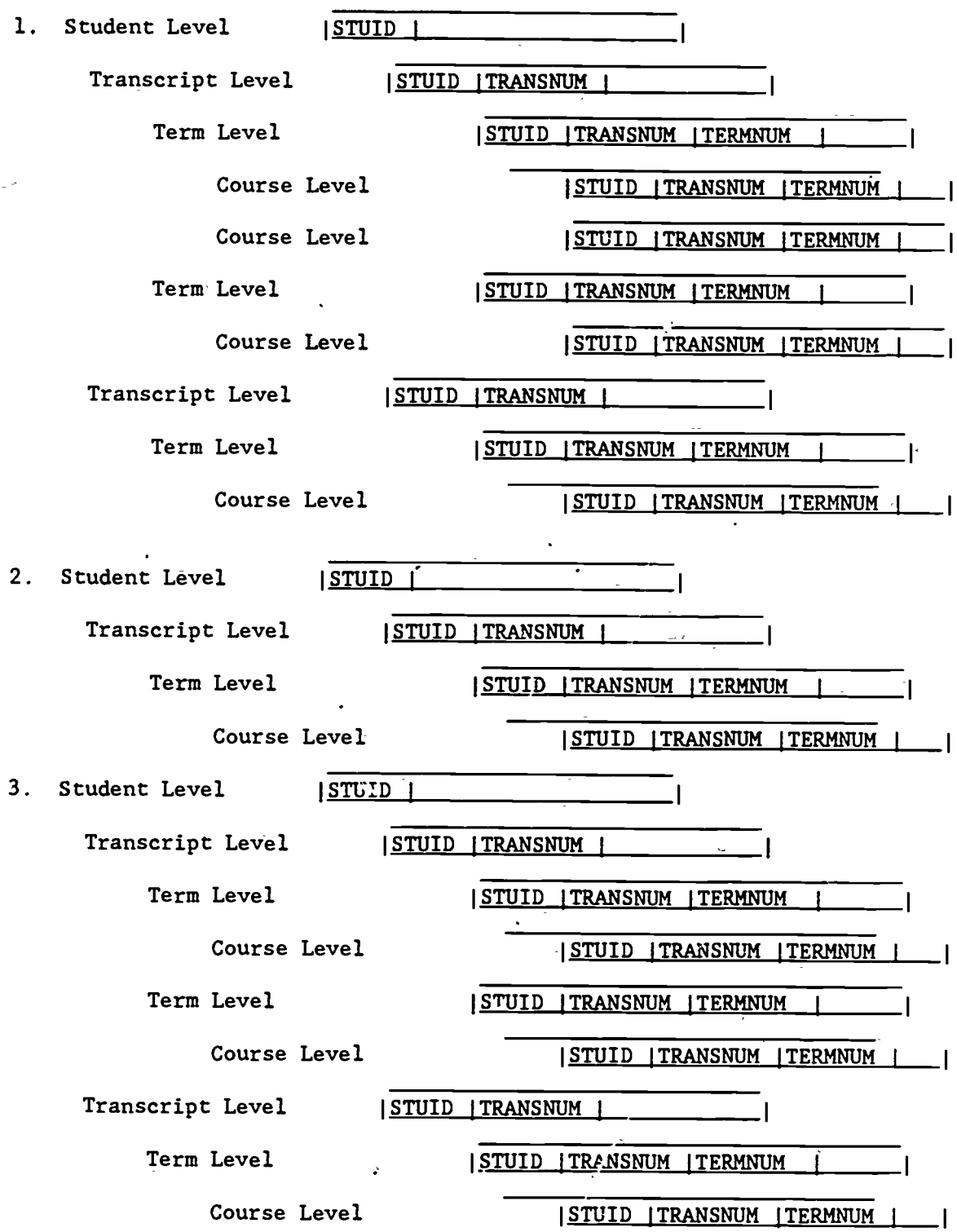
As we have emphasized throughout this report, postsecondary transcript data were abstracted from school records of greatly varying structure and content. It is essential for researchers using these data to be fully aware that the elements in the database are intended to be a faithful reproduction of the information reported on the transcripts. Except for the creation of limited composite variables, the transcript data have not been rescaled, standardized, or otherwise manipulated prior to entry into the database. For some items, notably course grades, school-reported grade point averages, and course credits, the researcher must not assume that the data stored in the designated fields are all values from a common underlying metric.

Course grades were entered as they appeared on the transcript. Two types of grades (letter and numeric) were stored on separate fields in the course records in order to minimize the effort needed to compute customized grade indicators.

As explained above, a comprehensive list of allowable letter grades (including such administrative "grades" as "credit given," "audit," "withdrawal," "pass," and "fail") was constructed to handle the entry of letter grades reported by schools. Although nearly all (97.3 percent) of the schools assigned letter grades, not all schools used all possible grades to make distinctions between student performance levels. Although most schools used conventional "+" and "-" qualifiers, some schools applied these only to selected levels (e.g., C+, C, and C-, but not B+ or B-). More important, however, is the fact that several different schemes of numeric equivalents were used by schools in translating letter grades to number grades for the computation of grade point averages. By far the most common scheme is the standard four-point collegiate scale (A - 4, B - 3, C - 2, D - 1, E or F - 0). A small number of schools assigned different numeric equivalents, however, such as setting the value of an "A" grade to 5 or 6

Exhibit 4-1

A Schematic Diagram of the Database Hierarchy
Representing Nested Transcript, Term, and Course Records
for Three Sampled Students



numeric points. For this reason, some grade point averages in the fields on the transcript-level records may exceed the conventional upper bound of 4.0.

Less that 3 percent of all courses in the file were graded on a numeric scale. These courses were disproportionately found on transcripts from short-term vocational/proprietary school programs. To help establish a basis for standardizing the metric for the numeric grades, the term records contain fields showing the highest, lowest, and minimum passing scores for the designated school's grading system if this information was present on the transcript or in other documentation (bulletins or course catalogs) from the school.

The data in the course credits field also were entered exactly as reported on the transcript form, with no attempt made to standardize the units. Researchers should use special caution in analyzing and further manipulating course credit data. At a minimum, researchers should familiarize themselves with the variability of data in the fields prior to conducting analyses. We further recommend that researchers carefully examine the ranges and distributions of credit values reported by different types of institutions. For the most part, standard collegiate institutions reported credits based on the same or very closely related credit scales. At these institutions, the typical academic course in most departments carried a value of 3 credits, and so this is the modal value observed for courses at these institutions. (In fact, 51 percent of all courses taken by HS&B sample members carried exactly 3 credits.) A significant proportion of courses, especially those in the hard sciences that included extensions such as laboratory periods and other additions to standard classroom schedules, earned higher credit values, although the majority of the values fall between 3 and 5 credits for these expanded courses. Lower-level courses whose classes met for fewer hours per week had credit values between 1 and 3 (about 24 percent of all courses taken).

Courses with credit values greater than 5 were rare (about 2 to 3 percent of all courses for which credits could be coded). Altogether, 95 percent of the courses taken by sample members carried between 0 and 5 credits (about 5.8 percent of the courses carried no credit). Courses with credit values between 5 and 20 accounted for an additional 1.5 percent of all courses taken by HS&B cases. Credit values greater than 20 and up to the allowable limit of 999.997 (almost exclusively from vocational programs reflecting clock-hour systems) accounted for less than 1 percent of those recorded.

For most conventional analyses, researchers may wish to record credit values above, for example, 5 or 8 credits, to the missing data code in order to prevent unusual programs with extreme values from affecting results. Researchers who are especially interested in vocational programs and courses should carefully examine all of the data related to courses with high credit values.

Of concern in any analysis of course credits is the possibility of differences between the numeric scales for credits awarded by schools on the semester system and schools on the quarter system. These two types of term systems accounted for about 79 percent of all terms in which sampled students were enrolled. Trimesters accounted for 16 percent of the 43,592 term records. There were fewer other types of terms in the transcripts. Variable-length terms,

common at vocational schools, accounted for an average of 1 percent of all terms reported. Semesters, on the other hand, accounted for 65 percent of the terms; quarters accounted for 13.5 percent.

Typically, the number of credits required for graduation from schools on the quarter system is slightly higher than the number required by schools on the semester system. This gives rise to the concern that course credits may not be expressed in comparable units across types of institutions, and that the value of a course given at a quarter system school may have "inflated" credit value, compared to the credit value of the same courses at a school on the semester system. Some researchers have suggested that the transcripts data file include additional fields containing rescaled or standardized credits, to ensure that credits from differing systems were scored on a common metric. A frequent suggestion has been that course credits for schools on the quarter system be deflated by a linear transformation in order to more nearly equal those awarded by semester system schools.

Although the Postsecondary Education Transcript Study did not include the resources for a formal study of this issue, a number of empirical analyses demonstrated that the fact of comparability or non-comparability cannot easily be established. The simplest but most compelling evidence against any simple transformation of quarter system course credits came from comparisons of the credit values of standard collegiate courses taken by students in both types of schools. These comparisons showed clearly that for most typical science, mathematics, social science, or humanities courses, the credit values were the same (generally 3 credits) at both types of institutions. Further comparisons of the average number of credits carried by students per terms showed no systematic or significant differences between the two systems. For these reasons, the final decision concerning course credits was to include on the public release tapes only the raw credit values as they were reported on the transcripts, and to caution researchers that the comparability of credits across institution and term types could not be assumed, but should be carefully assessed in light of specific analytical objectives.

Finally, a major source of variation in the credit values in the file relates to the use of "clock hours" rather than conventional "credit hours" by vocational and proprietary schools. Students at these schools often earned several hundred clock hour credits for completing a unified program made up of several instructional modules each lasting a few days. Analysts are strongly urged to use special caution in the analysis of course credit fields because of the extreme effects these outlier values (some ranging as high as 999.997) may have on statistical estimates. These values have been retained in the system to support special analyses of relatively small subgroups of students and their educational activities. Failure to provide for special handling of these cases may produce bizarre results in conventional analyses.

5. SAMPLE DESIGN AND IMPLEMENTATION

The Sophomore Postsecondary Education Transcripts Study involved the collection and processing of school records for a subsample of the High School and Beyond (HS&B) 1980 sophomore cohort. A full description of the sample design for HS&B is provided in the sample design reports for the base year and first, second, and third follow-up surveys.[1] The following sections present an overview of the sample design for the full survey.

5.1 Base Year Sample Design

The base year (1980) survey employed a two-stage, highly stratified sample design with secondary schools having tenth and/or twelfth grades as the first-stage units of selection and students within schools as the second-stage units. With the exception of certain special strata, which were oversampled, schools were selected with probabilities proportional to their estimated enrollment in the tenth and twelfth grades. Within each school, 36 seniors and 36 sophomores were randomly selected. In schools with fewer than 36 seniors or 36 sophomores, all eligible students were selected. Sampling rates were set so as to select within each stratum the number of schools needed to satisfy study design criteria regarding minimum sample sizes for certain types of schools. As a result, some schools had a very high probability of inclusion in the sample (in some cases equal to 1.0) while others had a much lower probability of inclusion. The total number of schools selected for the initial sample was 1,122, from a frame of 24,725 schools with grades ten or twelve or both.[2] Sampling strata and the number of schools selected in each are shown in Table 5-1.

Substitutions were made for schools that refused to participate in the survey. No substitutions were made, however, for students who for whatever reason failed to participate.[3] Substitutions for refusal schools occurred only within strata. In certain cases no substitution was possible because a school was the sole member of its stratum. (See the High School and Beyond Third Follow-Up Sample Design Report, which is available from NCES.)

The realization of the sample by stratum is shown in Table 5-2. Although the sample design specified that students in all but the special strata would be selected with approximately equal probabilities, the probabilities are only roughly equal. In the special strata, students were selected with higher probabilities--in some instances, extremely high probabilities. Moreover, the sample as realized did not equal the sample as drawn, creating further deviations from a self-weighing sample. Consequently, each school (and student) was assigned a weight equal to the number of schools (or students) in the universes they represented. Since each student's overall selection probability (hence weight) was further influenced by the sample design, the derivation of student case weights is discussed below. Calculation of school weights is described in the users' manual for the school questionnaire data file.

Table 5-1

High School and Beyond Base Year School Sample Selection

	Number
Special strata (oversampled)	
Alternative public	50
Cuban public	20*
Cuban Catholic	10*
Other Hispanic public	106*
High performance private	12
Other non-Catholic private (stratified by four Census regions)	38
Black Catholic	30*
Regular strata (not oversampled)	
Regular Catholic (stratified by four Census regions)	48
Regular public (stratified by nine Census divisions; racial composition; enrollment; central-city, suburban, rural)	808
	1,122

*These schools were defined as those having 30 percent or more of enrollment from the indicated subgroup.

Table 5-2

High School and Beyond Base Year Sample Realization

Stage 1: Sampling of schools

Stratum	Drawn in sample	Original schools*	Substituted schools	Total realized
Regular public	808	585	150	735
Alternative public	50	41	4	45
Cuban public	20	11	--	11
Other Hispanic public	106	72	30	102
Regular Catholic	48	40	5	45
Black Catholic	30	23	7	30
Cuban Catholic	10	7	2	9
High performance private	12	9	2	11
Other non-Catholic private	38	23	4	27
TOTAL	1,122	811	204	1,015

Stage 2: Sampling of students

	Total drawn in sample	Absent, both Survey and Make-Up Days	Student refused	Parent refused	Partial materials missing**	Total realized
Number	70,704	8,278	1,759	233	2,174	58,270
Percent	100	12	3	-	3	82

*Includes additional selections made when schools were found to be out-of-scope.
**Unusable because critical survey materials missing.

Use of appropriate weights should lead to correct estimates (within sampling error) of the population of tenth and twelfth grade students in United States schools in spring 1980, and of subgroups within that population.

5.2 1980 Sophomore Cohort Sample Design for Second and Third Follow-Up Surveys

The sample design for the 1980 sophomore cohort was based on the high school transcript study conducted between the first and second follow-ups. During the fall of 1982, high school transcripts were sought for a probability subsample of nearly 18,500 members of the 1980 sophomore cohort. The subsampling plan for the transcript study emphasized the retention of members of subgroups of special relevance for education policy analysis. Compared to the base year and first

follow-up, the transcript study sample design further increased the over representation of racial and ethnic minorities (especially those with above-average HS&B achievement test scores), students who attended private high schools, school dropouts, transfers and early graduates, and students whose parents participated in the base year parent survey on financing postsecondary education.

Transcripts were collected and processed for nearly 16,000 members of the sophomore cohort. A public use data file containing transcript information is available from NCES. Transcript data can be merged easily with student questionnaire data files using the case identification numbers common to the two files. The Data File Users Manual for the HS&B High School Transcripts Study (also available from NCES) contains a full description of the sample design and other features of the transcript study.

The sample for the second follow-up survey of the 1980 sophomore cohort was composed of approximately 15,000 cases selected from among the 18,500 retained for the transcript study. Like the second follow-up sample for the senior cohort, the sample for the sophomore cohort includes disproportionate numbers of persons from policy-relevant subpopulations--for example, racial and ethnic minorities, students from private high schools, high school dropouts, and students who planned to pursue some type of postsecondary schooling. The sample for the third follow-up survey was identical to that of the second follow-up. The second/third follow-up sample, though much smaller than the base year and first follow-up samples, is thus able to provide estimates for many subpopulations that are nearly as precise, statistically, as those of the larger samples. The second and third follow-up sample allocation is shown below in Table 5-3. For further details see the High School and Beyond Second Follow-Up Sample Design Report, by C. Jones and B. Spencer (NORC, 1984). The base year and first follow-up sample report is available from NCES.

5.3 The Senior Cohort Postsecondary Education Transcript Study (PETS) Sample

In 1984, postsecondary transcripts were requested for all members of the 1980 senior cohort who reported in either the first or second follow-up survey attending any form of postsecondary school since leaving high school. Thus, no further probabilistic sampling was done to define the PETS sample. The only restriction on inclusion in the PETS sample was that the respondent must have provided the name of the school attended, so that records could be requested. Thus, omitted from the transcript study were a very few sample members who indicated that they had attended some form of postsecondary school, but who gave no indication during either follow-up survey of the name of the school(s). In all, 7,776 members of the 1980 senior cohort satisfied the initial criteria for inclusion by naming at least one school in at least one of the follow-up surveys.

5.4 The Sophomore Cohort Postsecondary Education Transcript Study Sample

In order to conserve resources, a somewhat more restrictive sample was drawn for the HS&B sophomore cohort than was drawn for the senior cohort. The Department of Education was primarily interested in learning about the HS&B sample members who exhibited a "normal" pattern of postsecondary school attendance. Therefore, it was decided at the outset that those students who

Table 5-3

1980 Sophomore Cohort Second Follow-Up Sample
Distribution by Race-Ethnicity Typology

Student Status Category	Population size N	Population size % of Total	Second follow-up N	Second follow-up % of Total
Hispanic				
Cuban/Puerto Rican	89,674	2.4%	990	6.7%
High achievement	85,762	2.3%	886	6.0%
Other Hispanic	299,802	7.9%	1,375	9.3%
Asian/Pacific Islander	46,835	1.2%	431	2.9%
Native American	48,418	1.3%	291	2.0%
Black				
High achievement	84,544	2.2%	741	5.0%
Other	375,185	9.9%	1,295	8.7%
High Achievement/ Low-SES whites	69,759	1.8%	388	2.6%
All others	2,679,309	70.9%	8,428	56.8%
TOTAL	3,779,288	100.0%	14,825	100.0%

NOTE: For this typology, sample members were assigned to ethnic or racial categories on a sequential or hierarchical basis. That is, individuals who reported Cuban or Puerto Rican origin or descent in either the base year or first follow-up were so classified in this typology. High-achievement Hispanics were then classified among the remaining non-Cuban/non-Puerto Rican cases. (Since some Cubans and Puerto Ricans were also "high achievement," the total number of high-achievement Hispanics is larger than shown in this table. "Other Hispanics" were then classified from among all remaining cases not assigned to the two previous categories. This procedure was repeated sequentially for each remaining category in the table. The result is a distribution of mutually exclusive categories whose contents sum to the population or sample size. The distributions presented mask considerable overlap among groups within the sample (e.g., Blacks who are also Hispanic).

entered postsecondary school in the fall immediately following their high school graduation would be drawn into the sample. With the exception of vocational students, students who delayed their postsecondary school until the winter of 1983 or later were not included in the sample.

No probabilistic sampling was undertaken; rather, students who were considered of greatest policy interest were selected into the sample with certainty. More specifically, the sample was selected in two steps. First, students exhibiting certain attendance patterns were selected, and second, the schools they attended were selected. Students were selected into the sample on the basis of their responses to second follow-up (1984) and third follow-up (1986) questions on schools attended after leaving high school.

Under Step 1, students defined as normal persisters were drawn into the sample with certainty. Normal persisters were students who began attending any postsecondary school (with the exception of foreign schools) full-time by October 1982 and did not leave the school until after August of 1982. This definition removes students who attended school during the summer only in 1982. Normal persisters attended any of six types of schools: proprietary, private technical or two-year, public technical, two-year college or university, four-year public university, and four-year private college or university.

Next, vocational students were drawn into the sample. These students were not normal persisters and started attending a proprietary school, private technical or two-year school, or public technical school and did not leave until after August 1982. Again, this definition eliminates students who studied in the summer only. Vocational students were included even if they were attending school part-time.

Under Step 2, the schools were selected. Because a certain proportion of students transfer from their first school to other schools, there are necessarily more transcripts than students. In fact, the sample that results from the two-step selection process is a sample of student-school combinations for which transcript information is collected.

No attempt was made to request transfers from all schools attended by the sample of students. Transcripts were selected from second and third schools only if they represented a pattern of normal progression through postsecondary school. The schools were selected as follows:

- If a student was a normal persister and started attending a two-year public, four-year private, or four-year public college or university, this school was selected.
- Any other four-year private or four-year public institution was selected if, after attending the first school, the student began attending this school as a full-time student.
- Any two-year public university was selected if, after attending the first school, the student attended this school and also attended another four-year private or four-year public university.

- If the student was a normal persister and began attending a proprietary, private technical, or two-year school, or public technical school, this school was selected.

- If a student was a vocational student, then the first school was selected for this student. If a vocational student began attending a second vocational school, this school was not selected.

A total of 6,098 students and 2,139 schools were selected into the sample. Table 5-4 shows the distribution of students and transcripts.

However, there were 565 students for whom transcripts were not received. There were a variety of reasons given for not sending transcripts: schools refused to release transcripts; schools refused to cooperate with the transcript study; transcripts were lost or destroyed; schools closed; and there was no response from the school. In addition, there were some students whom school officials claimed had never enrolled or did not complete sufficient work to have an enrollment record.

Because the evidence for non-attendance is not completely conclusive for the students who reportedly "never attended" or any of the rest of the 565 cases, these students have been included on the public release data files (including raw weights and selected HS&B questionnaire data). These cases also have a single dummy Transcript Record whose Final Disposition field indicates the reason for no response. In the course of normal transcript data analysis, these cases may be deleted from the analysis data files by selecting for the analysis only cases with non-zero values for one of the transcript weights.

Table 5-4

High School and Beyond Sophomore Postsecondary Transcript Sample

Student Group	Students	Transcripts
Normal persisters in public 2-year, private 4-year, and public 4-year	5,122	6,453
Normal persisters in proprietary private technical 2-year, or public technical school	572	572
Vocational students	404	404
TOTAL	6,098	7,429

Table 5-5 shows the distribution of the number of schools reported by students who were considered in-scope. The only students considered out of scope were those 229 who reportedly "never attended" the institutions they had named in the second or third follow-up survey. The analyst will note that the transcript level file is coded "never attended". Also deleted from this table are 47 duplicate transcripts. Over three-fifths of the students reported attending only one institution in their responses to the follow-up surveys. An additional 30 percent of these cases reported attending exactly two schools. Only about 8 percent (602) reported attending three or more postsecondary schools during the four-year period since leaving high school.

5.5 Sample Weights

The general purpose of weighing survey data is threefold: the weights allow data from the sample to be used for estimating population totals; the weights compensate for unequal probabilities of selection (or retention) in the survey; and the weights adjust for nonresponse in the study.

The HS&B weights are based on the inverse of the selection probabilities through all stages of the sampling process; the nonresponse adjustments are based on the inverse of the response rates within weighing classes. A "raw" weight, which reflects only the selection probabilities and which is not adjusted for nonresponse, is also calculated and will be included on the data files for the Postsecondary Transcript Study. The raw weight allows analysts to construct their own adjustments for nonresponse; in addition, the raw weight was used in calculating weighted response rates for the purpose of nonresponse adjustment.

Table 5-5

Number of Postsecondary Schools Reported by Members of the HS&B 1980 Sophomore Cohort

Number of schools	Number of cases	Percent
One	4,606	62.0 %
Two	2,226	29.9 %
Three or more	602	8.1 %
TOTAL*	7,434	100.0 %

*NOTE: An additional 342 cases who reported attending a single school were defined as ineligible and are excluded from this table.

The weighing procedures for the Postsecondary Education (PSE) Transcripts Study involved two major steps:

Step 1. Calculation of a preliminary, or raw, weight based on the inverse of the product of the probabilities of selection for the base year sample and retention in the follow-up surveys. This new raw weight is simply the follow-up raw weight times the inverse of the probability of retention in the PSE Transcript sample.

Step 2. Adjustment of the raw weight to compensate for "unit" nonresponse--that is, for nonresponse on an entire questionnaire, test, or transcript. (By definition, the new raw weight, RAWWT, is unadjusted for nonresponse.)

For the sophomore cohort, the PSE Transcript Study involved no new subsampling beyond what had been carried out for the HS&B second follow-up; that is, all second follow-up cases deemed eligible for the PSE Transcript Study were included in the sample. (Relative to the senior cohort, a somewhat more restrictive definition of eligibility was used in designating cases for the sophomore cohort PSE Transcript sample. The sample consisted mainly of students who enrolled full-time in fall 1982 in an academic institution or who attended a vocational technical school any time before July 1986). Thus, the raw weight described in Step 1 above is the same as the raw weight for the second (and third) follow-up survey.

Two separate nonresponse adjustments were calculated using the general technique described in Step 2. Both sets of nonresponse adjustments apply to all 6,098 cases selected for the PSE Transcript Study. The first adjustment corrects for nonresponse in the Transcript Study itself. For the purpose of this adjustment, a case was counted as complete if one on more transcripts were obtained for that case; a case was treated as a nonrespondent if no transcripts were obtained. The second adjustment corrects for nonresponse in the Transcript Study and the four prior surveys (i.e., the base year and three follow-ups). For the purpose of this adjustment, a case was counted as complete only if the case had at least one transcript and completed questionnaires for all four HS&B survey rounds; all other cases were counted as nonrespondents.

This approach to weighing defines the sample person, rather than the individual transcript, as the unit of analysis. The weights apply to the person and to all the data associated with that person and are not intended to be applied to individual transcripts.

Both sets of nonresponse adjustments were computed as simple ratios (sum of the raw weights for all cases over the sum of the raw weights for the completed cases) within 29 weighing cells. The weighing cells were defined by cross-classifying cases according to the type of high school attended, sex, race, and the type of postsecondary school attended. These four variables have been consistently related to nonresponse in the HS&B studies and were used in defining nonresponse adjustment cells for the Senior PSE Transcript Study. The cross-classification results in 48 cells; cells with 20 or fewer cases were pooled with adjacent cells having similar completion rates. (In a few cases, small cells

requiring nonresponse adjustments close to 1.0 were left intact.) After pooling, 29 cells remained; nonresponse adjustments were calculated for each cell.

Within each cell, the nonresponse adjustment was obtained by dividing the sum of weights for all selected cases by the sum of weights for the "completed" cases. The nonresponse adjustment is thus the inverse of the weighted response rate. The final adjusted weights are just the product of the adjustment factors and the raw weights. Tables 5-6 and 5-7 below present the weighing cells and adjustment factors for both sets of PSE Transcript weights.

If a completed case is defined as one for which at least one transcript was obtained, then the weighted completion rate for the PSE Transcripts Study is 90.8 percent (1,292,191 weighted completes over 1,422,340 eligible; see Table 5-6). The average adjustment factor is just the inverse of this completion rate (i.e., 1.10). Similarly, if a completed case is defined as one with at least one transcript and questionnaire data from all prior waves of the survey, the weighted completion rate is 79.3 percent (see Table 5-7), and the mean adjustment factor is 1.26.

Relative to the senior cohort PSE Transcript weights, three differences are readily apparent. First, the size of the population for the sophomore cohort (estimated by the sum of the weights) is smaller than that for the seniors (1.4 million versus 1.8 million; cf. Tables 5-6 and 5-7 with Tables 5.4-1 and 5.4-2 in the High School and Beyond Senior Cohort Postsecondary Education Transcript Study Data File Users' Manual). This difference appears to reflect the more restrictive criteria used in defining eligibility for the PSE Transcript Study within the sophomore cohort.

A second difference is that the adjustment factors are somewhat larger for the sophomore cohort than for the senior cohort; this reflects the difference in response rates. Overall, at least one transcript was obtained from about 94 percent of the senior cohort sample (versus 91 percent for the sophomore cohort). Similarly, cases with at least one transcript and complete questionnaire data from prior rounds constituted 86 percent of the senior cohort PSE Transcript sample (versus 79 percent for the sophomore cohort).

Finally, we note that for most of the weighing cells involving cases with only vocational postsecondary education (rows 23 through 29 in Tables 5-6 and 5-7), the estimated population sizes are actually somewhat larger for the sophomore cohort than for the senior cohort, despite the reduction in the overall population size noted earlier (cf. Tables 5.4-1 and 5.4-2 in the senior cohort Data File Users' Manual). This appears to reflect a real underlying difference between the two cohorts and is consistent with other data from the High School and Beyond surveys. For example, as of the third follow-up (when cases were selected for the sophomore PSE Transcripts Study), about 15 percent of the sophomore cohort reported that they had attended vocational school; the corresponding figure for the senior cohort (as of the second follow-up, when cases were selected for the senior PSE transcript study) is only 11 percent.

Table 5-6

Nonresponse Adjustments to Sampling Weights
for Completed Cases in HS&B Sophomore Cohort
Postsecondary Education Transcript Study (WT1)

Weighing classes

	Vocational postsecondary only	Type of secondary school	Sex	Race	Sum of weights: eligible	Sum of weights: completes	Nonresponse adjustment
1	No	Reg Public	M	Hisp	17,799	16,848	1.0565
2	No	Reg Public	M	Black	35,482	31,721	1.1185
3	No	Reg Public	M	Other	375,958	360,566	1.0426
4	No	Reg Public	F	Hisp	14,856	12,957	1.1465
5	No	Reg Public	F	Black	66,612	63,151	1.0548
6	No	Reg Public	F	Other	445,419	427,335	1.0423
7	No	Hisp Public	M	Hisp	5,516	4,744	1.1625
8	No	Hisp Public	M	Black	800	753	1.0614
9	No	Hisp Public	M	Other	4,848	4,688	1.0341
10	No	Hisp Public	F	Hisp	6,579	6,370	1.0328
11	No	Hisp Public	F	Black	2,608	2,172	1.2009
12	No	Hisp Public	F	Other	5,227	4,791	1.0911
13	No	Catholic	M	Hisp	1,924	1,733	1.1104
14	No	Catholic	M	Black	2,138	2,014	1.0612
15	No	Catholic	M	Other	53,256	50,106	1.0629
16	No	Catholic	F	Hisp	4,180	4,064	1.0285
17	No	Catholic	F	Black	4,770	4,264	1.1186
18	No	Catholic	F	Other	55,698	53,106	1.0487
19	No	Oth Private	M	Other	30,138	27,855	1.0820
20	No	Oth Private	F	Other	32,249	31,251	1.0319
21	No	Oth Private	M&F	Hisp	1,631	1,631	1.0000
22	No	Oth Private	M&F	Black	2,010	2,006	1.0020
23	Yes	Reg Public	M	Hisp	6,714	2,517	2.6673
24	Yes	Reg Public	M	Black	12,019	8,301	1.4478
25	Yes	Reg Public	M	Other	73,211	55,113	1.3283
26	Yes	Reg Public	F	Hisp	10,314	4,218	2.4449
27	Yes	Reg Public	F	Black	24,429	12,872	1.8978
28	Yes	Reg Public	F	Other	97,540	72,065	1.3535
29	Yes	All Private	M&F	All	28,416	22,979	1.2365
	TOTAL				1,422,340	1,292,191	

Table 5-7

Nonresponse Adjustments to Sampling Weights
for Cases with At Least One Postsecondary Transcript and
Completed Questionnaires from the Base Year, First, Second, and Third
Follow-Up Surveys (WT2)

Weighing Classes

	Vocational postsecondary only	Type of secondary school	Sex	Race	Sum of weights: Eligible	Sum of weights: Completes	Nonresponse adjustment
1	No	Reg Public	M	Hisp	17,799	13,930	1.2777
2	No	Reg Public	M	Black	35,482	27,538	1.2884
3	No	Reg Public	M	Other	375,958	315,076	1.1931
4	No	Reg Public	F	Hisp	14,856	11,569	1.2840
5	No	Reg Public	F	Black	66,612	51,900	1.2835
6	No	Reg Public	F	Other	445,519	381,640	1.1671
7	No	Hisp Public	M	Hisp	5,516	3,519	1.5676
8	No	Hisp Public	M	Black	800	729	1.0968
9	No	Hisp Public	M	Other	4,848	3,761	1.2889
10	No	Hisp Public	F	Hisp	6,579	5,189	1.2679
11	No	Hisp Public	F	Black	2,608	2,172	1.2009
12	No	Hisp Public	F	Other	5,227	4,733	1.1044
13	No	Catholic	M	Hisp	1,924	1,676	1.1479
14	No	Catholic	M	Black	2,138	1,696	1.2600
15	No	Catholic	M	Other	53,256	46,264	1.1511
16	No	Catholic	F	Hisp	4,180	3,901	1.0713
17	No	Catholic	F	Black	4,770	3,884	1.2281
18	No	Catholic	F	Other	55,698	50,526	1.1022
19	No	Oth Private	M	Other	30,138	22,389	1.3460
20	No	Oth Private	F	Other	32,249	24,516	1.3154
21	No	Oth Private	M&F	Hisp	1,631	1,251	1.3044
22	No	Oth Private	M&F	Black	2,010	1,829	1.0989
23	Yes	Reg Public	M	Hisp	6,714	2,117	3.1715
24	Yes	Reg Public	M	Black	12,019	6,959	1.7271
25	Yes	Reg Public	M	Other	73,211	44,991	1.6273
26	Yes	Reg Public	F	Hisp	10,314	2,393	4.3103
27	Yes	Reg Public	F	Black	24,429	10,066	2.4265
28	Yes	Reg Public	F	Other	97,540	61,099	1.5964
29	Yes	All Private	M&F	All	28,416	20,642	1.3766
	TOTAL				1,422,340	1,127,955	

Table 5-8 shows the statistical properties of the raw weights and the two sets of adjusted weights. The table includes the mean, sum, variance, standard deviation, coefficient of variation, minimum, maximum, skewness, kurtosis, and the number of weighted cases for each weight. Note that each of three weights is constrained to sum to the same estimated population total (1,442,340). Similarly, the sums of the three weights are constrained to be equal within each of the 29 weighing cells.

5.6 Standard Errors and Design Effects

Statistical estimates based upon High School and Beyond data are subject to sampling variability. Sampling errors arise because data are collected from only a randomly selected portion of the members of a population of interest. The HS&B sophomore cohort sample, as realized, is only one representation of a large number of samples of similar size that might have been drawn. Sampling errors are directly related to the underlying variability of the property being measured, and are inversely related to the number of observations contributing to the statistical estimates.

Because the sample design for the HS&B cohorts involved stratification, disproportionate sampling from certain strata, and clustered (i.e., multi-stage) probability sampling, the calculation of exact standard errors for survey estimates can be difficult and

Table 5-8

High School and Beyond Sophomore Cohort
Postsecondary Education Transcripts Study
Statistical Properties of Sample Case Weights*

Weight	RAWWT	WT1	WT2
Mean	233	257	289
Sum	1,422,340	1,422,340	1,422,340
Variance	42,967	53,930	68,608
Standard deviation	207	232	262
Coefficient of variation	89	90	91
Minimum	1	1	1
Maximum	2,219	2,392	3,058
Skewness	1.65	1.77	1.78
Kurtosis	7.1	7.7	8.2
Number of cases	6,098	5,533	4,930

*NOTE: All entries except skewness and kurtosis have been rounded to the nearest whole number; the coefficient of variation is in percentage terms.

expensive. Popular statistical analysis packages such as SPSS (Statistical Programs for the Social Sciences) or SAS (Statistical Analysis System) normally calculate standard errors using the assumption that the data being analyzed were collected from simple random samples. As is described in detail in the High School and Beyond sample design reports for each survey wave, the HS&B sample design is, on balance, somewhat less efficient than simple random samples of equal size. Thus, sampling errors generated by SPSS and SAS will normally underestimate significantly the sampling variability of statistical estimates such as population means, percentages, and more complex statistics such as correlation and regression coefficients.

Several procedures are available for calculating precise estimates of sampling error for complex samples. Procedures such as Taylor series approximations, Balanced Repeated Replication (BRR), and Jackknife Repeated Replication (JRR) vary somewhat in computational convenience and cost, and in their ability to account for several sources of sampling variability, most notably clustered selection of sample cases.

After each survey wave since the base year, sampling variances have been calculated for about thirty estimated proportions or means for the whole sample and for several subgroups (domains), and have been reported in the data file users' manuals for each public release tape. In general, these calculations have been carried out using BRR. However, comparisons of variance estimates provided by Taylor series and BRR carried out at the time of the HS&B first follow-up survey showed little difference in the resulting error estimates for such statistics as means, proportions, and Pearson correlation coefficients.

In addition to standard errors, the design effects for each estimate (DEFF) and the square roots of each design effect (DEFT) were calculated and reported. The design effect is a measure of the inefficiency of the sample estimate relative to a simple random sample of equal size. It is defined as the ratio of the actual variance of an estimate (i.e., the square of the standard error) to the variance of the same estimate from a simple random sample with the same number of cases. For proportions, the estimated simple random sample variance is just

$$VAR(SRS) = p(1 - p)/n \qquad (1)$$

in which

p = the estimated proportion

and

n = the number of cases with non-missing data

Like almost all national samples, the High School and Beyond sample is not a simple random sample. The High School and Beyond sample departs from the model of simple random sampling in three major respects: the observations are clustered at the school level; major groups (such as students who attended private schools) are deliberately represented disproportionately; and the sample is stratified by type of school. Each of these departures from simple random sampling has an effect on efficiency, which is reflected in the design effect.

Separate sampling errors and design effects have not been calculated for the postsecondary transcript data. The calculations of sampling errors and design effects performed for the High School and Beyond 1980 Sophomore Cohort Second Follow-Up (1984) Data File Users Manual have been reproduced and included in Table 5-9.

The mean design effects given in Table 5-9 can be used to calculate approximate standard errors for estimates based upon transcript data. For example, the standard error of a proportion can be estimated using the square root of the expression in (1) (above) times the mean root design effect (DEFT):

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

With the exception of those for Hispanics, the DEFTs in Table 5-6 for subgroups are generally 10 percent smaller than that for the total population. The relative efficiency of the Hispanic subsample continues to be affected by the somewhat larger follow-up cluster sizes for Hispanic sample members in specific schools and relatively few geographical areas, and higher variability in sample weights because some Hispanics (those in so-called "Hispanic schools") were sampled at very high rates while others (in regular public schools) were sampled at rates closer to those of majority whites. Furthermore, the variability of the DEFTs for Hispanics is over twice that observed for most other subgroups. Thus, for analysis of data from Hispanics, the use of a single generalized design effect to inflate simple random sample estimates of sampling errors involves a larger degree of DEFTs for approximation. Nevertheless, the differences between Hispanics and other groups remain generally small. Researchers who use design effect factors to estimate standard errors for Hispanic sample data and who prefer to be statistically conservative may wish to choose a design effect slightly larger than the mean of 1.48 in Table 5-9.

In addition, Table 5-10 presents selected distributional statistics for the DEFF and DEFT factors for proportions taken from prior survey waves. These tables as well as several informal analyses carried out at NORC and at NCES, generally confirm that, with minor exceptions noted, the design effects have remained reasonably constant across survey waves and population domains, and show relatively small variability across survey items within waves and domains.

Table 5-9
Distributional Statistics for Design Effects and Root Design Effects for 30 Survey Measures for 12 Domains

Domain		DEFF	DEFT
Total population	Mean	2.19	1.48
	Minimum	1.40	1.18
	Maximum	2.68	1.64
	Standard deviation	0.29	0.10
Hispanic	Mean	3.11	1.75
	Minimum	1.69	1.30
	Maximum	5.40	2.32
	Standard deviation	0.76	0.21
Black	Mean	2.19	1.47
	Minimum	1.24	1.11
	Maximum	2.92	1.71
	Standard deviation	0.36	0.13
Whites and others	Mean	1.92	1.38
	Minimum	1.32	1.15
	Maximum	2.38	1.54
	Standard deviation	0.23	0.08
Female	Mean	2.06	1.43
	Minimum	1.51	1.23
	Maximum	2.42	1.55
	Standard deviation	0.21	0.07
Male	Mean	2.07	1.44
	Minimum	1.37	1.17
	Maximum	2.59	1.61
	Standard deviation	0.24	0.09
Lowest quartile SES	Mean	1.83	1.35
	Minimum	1.22	1.10
	Maximum	2.31	1.52
	Standard deviation	0.26	0.10
Middle quartiles SES	Mean	2.06	1.43
	Minimum	1.43	1.20
	Maximum	2.41	1.55
	Standard deviation	0.25	0.09
Highest quartile SES	Mean	1.92	1.38
	Minimum	1.31	1.14
	Maximum	2.48	1.57
	Standard deviation	0.28	0.10
Received no PSE	Mean	1.98	1.40
	Minimum	1.25	1.12
	Maximum	2.82	1.68
	Standard deviation	0.34	0.12
Received some PSE	Mean	2.09	1.44
	Minimum	1.46	1.21
	Maximum	2.53	1.59
	Standard deviation	0.19	0.07
Four-year degree	Mean	1.63	1.26
	Minimum	0.16	0.39
	Maximum	2.14	1.46
	Standard deviation	0.42	0.21

Table 5-10

Distributional Statistics for Design Effects and Root Design Effects for Proportions from Various Survey Waves
HS&B Sophomore Cohort

Survey	DEFF	DEFT
First Follow-Up, using First Follow-Up Weight		
Mean	3.14	1.72
Minimum	1.33	1.15
Maximum	7.41	2.72
Standard deviation	1.80	0.47
Changes in Proportions between BY and FFU, using FFU Weight		
Mean	1.80	1.33
Minimum	.95	.98
Maximum	3.45	1.86
Standard deviation	.61	.21
Second Follow-Up, using Second Follow-Up Weight		
Mean	2.40	1.54
Minimum	1.23	1.11
Maximum	4.00	2.00
Standard deviation	0.56	0.18

NOTES TO CHAPTER 5

[1]For further details on the base year sample design see Martin R. Frankel, Luane Kohnke, David Buonanno, and Roger Tourangeau, Sample Design Report (Chicago: NORC, 1981).

[2]The sampling frame, defined as the universe of high schools in the United States, was obtained from the 1978 list of U.S. elementary and secondary schools of the Curriculum Information Center, a private firm. This was supplemented by the NCES lists of public and private elementary and secondary schools. Any school listed in any of these files that contained a tenth grade, a twelfth grade, or both was made part of the frame.

[3]Apart from substitution for schools that refused, there were a number of schools in the originally-drawn sample that were "out of scope," that is, they failed to fit the criteria for inclusion in the sample. The sample was then augmented through selection of an additional school for each out-of-scope school, within major strata. Most of the out-of-scope schools were area vocational schools having no enrollment of their own, although they were listed in the frame as having enrollments.

Appendix A: List of Endorsing Institutions
Contents of School Transcript Request Packages

NATIONAL LONGITUDINAL STUDIES PROGRAM

High School and Beyond
A National Longitudinal Study for the 1980's

Sponsored by the Center for Education Statistics,
U.S. Department of Education

The professional organizations listed below fully endorse the Postsecondary Education Transcript Study and encourage their members to cooperate in this important project.

American Association of Collegiate Registrars and Admissions Officers (AACRAO)

American Association of Community and Junior Colleges (AACJC)
American Association of State Colleges and Universities (AASCU)
American Council on Education (ACE)
Association of Catholic Colleges and Universities (ACCU)
Association of Independent Colleges and Schools (AICS)
Association of Jesuit Colleges and Universities (AJCU)
The College Board
National Accrediting Commission of Cosmetology Arts and Sciences (NACCAS)
National Association of College and University Business Officers (NACUBO)
National Association for Equal Opportunity in Higher Education (NAFEO)
National Association of Student Financial Aid Administrators (NASFAA)
National Association of Trade and Technical Schools (NATTS)
National Council of Higher Education Loan Programs (NCHELP)
National Institute of Independent Colleges and Universities (NIICU)
United Negro College Fund, Inc.

UNITED STATES DEPARTMENT OF EDUCATION

OFFICE OF THE ASSISTANT SECRETARY
FOR EDUCATIONAL RESEARCH AND IMPROVEMENT

CENTER FOR EDUCATION STATISTICS

Dear Registrars and Officials:

As part of its Longitudinal Studies program, the Center for Education Statistics has been collecting transcript and other information for persons who have participated in it's surveys. To continue this effort, the Center has authorized the National Opinion Research Center (NORC) to obtain student transcript data for individuals who are participating in the High School and Beyond (HS&B) survey. The goal of this study is to provide information which can be aggregated to examine research issues at the national level. Education researchers and policy analysts will relate the information about courses taken and credits earned to the characteristics gathered from questionnaires and other sources. HS&B will enable researchers to analyze the relationships between coursetaking patterns, academic achievement, and subsequent occupational choices and success. Student names are used only to make sure that data on variables from different sources (tests, questionnaires, and transcripts) refer to the same individuals and not to find out anything about particular individuals.

The grant of authority for collection of the transcript data is made pursuant to the provision in the Family Education Rights and Privacy Act (FERPA) (20 U.S.C 1232g), implemented by 34 CFR 99.31(a)(6), that allows the release of records to the Secretary of Education or to his agent without the prior consent of the survey participants. The privacy of the information you are asked to supply to NORC will be protected, as required by FERPA. A copy of the relevant section of the act is reproduced on the reverse side of this page.

We would appreciate your cooperation with NORC in the transcript study.

Sincerely yours,

Emerson J. Elliott
Director

WASHINGTON, D.C. 20208

§ 99.31 Prior consent for disclosure not required.

(a) An educational agency or institution may disclose personally identifiable information from the education records of a student without the written consent of the parent of the student or the eligible student if the disclosure is—

(1) To other school officials, including teachers, within the educational institution or local educational agency who have been determined by the agency or institution to have legitimate educational interests;

(2) To officials of another school or school system in which the student seeks or intends to enroll, subject to the requirements set forth in § 99.34;

(3) Subject to the conditions set forth in § 99.35, to authorized representatives of:

(i) The Comptroller General of the United States,

(ii) The Secretary, or

(iii) State educational authorities;

(4) In connection with financial aid for which a student has applied or which a student has received; *Provided*, That personally identifiable information from the education records of the student may be disclosed only as may be necessary for such purposes as:

(i) To determine the eligibility of the student for financial aid,

(ii) To determine the amount of the financial aid,

(iii) To determine the conditions which will be imposed regarding the financial aid, or

(iv) To enforce the terms or conditions of the financial aid;

(5) To State and local officials or authorities to whom information is specifically required to be reported or disclosed pursuant to State statute adopted prior to November 19, 1974. This paragraph applies only to statutes which require that specific information be disclosed to State or local officials and does not apply to statutes which permit but do not require disclosure. Nothing in this paragraph shall prevent a State from further limiting the number or type of State or local officials to whom disclosures are made under this paragraph;

(6) To organizations conducting studies for, or on behalf of, educational agencies or institutions for the purpose of developing, validating, or administering predictive tests, administering student aid programs, and improving instruction; *Provided*, That the studies are conducted in a manner which will not permit the personal identification of students and their parents by individuals other than representatives of the organization and the information will be destroyed when no longer needed for the purposes for which the study was conducted; the term "organizations" includes, but is not limited to, Federal, State and local agencies, and independent organizations;

(7) To accrediting organizations in order to carry out their accrediting functions;

(8) To parents of a dependent student, as defined in section 152 of the Internal Revenue Code of 1954;

(9) To comply with a judicial order or lawfully issued subpoena; *Provided*, That the educational agency or institution makes a reasonable effort to notify the parent of the student or the eligible student of the order or subpoena in advance of compliance therewith; and

(10) To appropriate parties in a health or safety emergency subject to the conditions set forth in § 99.36.

(b) This section shall not be construed to require or preclude disclosure of any personally identifiable information from the education records of a student by an educational agency or institution to the parties set forth in paragraph (a) of this section.

(20 U.S.C. 1232g(b)(1))

§ 99.32 Record of requests and disclosures required to be maintained.

(a) An educational agency or institution shall for each request for and each disclosure of personally identifiable information from the education records of a student, maintain a record kept with the education records of the student which indicates:

(1) The parties who have requested or obtained personally identifiable information from the education records of the student, and

(2) The legitimate interests these parties had in requesting or obtaining the information.

(b) Paragraph (a) of this section does not apply:

(1) To requests by or disclosure to a parent of a student or an eligible student;

(2) To requests by or disclosures to school officials under § 99.31(a)(1);

(3) If there is written consent of a parent of a student or an eligible student, or

(4) To requests for or disclosure of directory information under § 99.37.

(c) The record of requests and disclosures may be inspected:

(1) By the parent of the student or the eligible student,

(2) By the school official and his or her assistants who are responsible for the custody of the records, and

(3) For the purpose of auditing the recordkeeping procedures of the educational agency or institution by the parties authorized in, and under the conditions set forth in § 99.31(a) (1) and (3).

(20 U.S.C. 1232g(b)(4)(A))

§ 99.33 Limitation on redisclosure.

(a) An educational agency or institution may disclose personally identifiable information from the education records of a student only on the condition that the party to whom the information is disclosed will not disclose the information to any other party without the prior written consent of the parent of the student or the eligible student, except that the personally identifiable information which is disclosed to an institution, agency or organization may be used by its officers, employees and agents, but only for the purposes for which the disclosure was made.

(b) Paragraph (a) of this section does not preclude an agency or institution from disclosing personally identifiable information under § 99.31 with the understanding that the information will be redisclosed to other parties under that section; *Provided*, That the recordkeeping requirements of § 99.32 are met with respect to each of those parties.

(c) An educational agency or institution shall, except for the disclosure of directory information under § 99.37, inform the party to whom a disclosure is made of the requirement set forth in paragraph (a) of this section.

(20 U.S.C. 1232g(b)(4)(B))

NORC

A Social Science Research Center
University of Chicago

1155 East 60th Street, Chicago, IL 60637
312/702-1200

June 1987

Dear Registrar:

NORC, a social science research center at the University of Chicago, requests your assistance in the conduct of a Postsecondary Education Transcript Study. We seek your help in collecting transcripts for a sample of students who are participating in the National Longitudinal Studies (NLS) program sponsored by the Center for Education Statistics (CES). The purpose of the transcript study, a component of NLS, is to obtain reliable and objective information about the types and patterns of courses taken by students. The data will make it possible for researchers to relate course-taking patterns to student characteristics available in student questionnaire files, and to subsequent occupational choice and success.

The NLS includes the National Longitudinal Study of the High School Class of 1972 (NLS-72) and of High School and Beyond (HS&B), the latter conducted by NORC since 1979. NLS-72 and HS&B constitute a large-scale, longitudinal study of the high school classes of 1972, 1980, and 1982. Nationally representative samples of the class of 1972 have been resurveyed five times since graduation, and the classes of 1980 and 1982 at two-year intervals since 1980. Approximately 16,000 members of the class of 1982 have reported attending about 2,100 postsecondary schools.

We would like to obtain the transcripts of one or more sample members who reported attending your school. Specifically we are requesting photo-copies of transcripts for each individual named on the enclosed checklist for the years reported by the student for his or her attendance. We would also appreciate it if you could provide us with: 1) a copy of the school's course catalog and 2) an interpretation of your grading system in order to facilitate accurate and uniform coding of the data. The folder contains more information about the study and our request for data. You will also find materials concerning applicable federal regulations and endorsements by professional organizations.

Privacy and confidentiality are always of concern to institutions and offices that maintain student records. CES and the organizations under contract to it adhere to the highest standards in protecting the privacy of individuals involved in the research it undertakes. Appropriate measures are employed to ensure the confidentiality of research participants during the collection, analysis, and reporting of all survey data. Of course, all relevant safeguards will be applied to this study.

Data are being collected under the provision of the Family Education Rights and Privacy Act (FERPA) that allows the release of records to the Secretary of Education or his agent without prior written consent by survey subjects. The same provision, 34 CFR 99.31 (6), applied to NORC's recent collection of high school transcripts of some 18,500 sample members and the collection of postsecondary school financial aid records for some 15,000 members. Both the purpose for and the manner in which the transcript data is to be acquired are in keeping with the FERPA requirements.

Endorsement of the transcript study has been made by the organizations listed on the cover of the folder. They welcome any inquiry you may wish to make regarding their support of the study.

We would appreciate return of the requested materials by August 1, or as soon thereafter as possible. Reimbursement for all transcripts will be made if you request it, and a voucher has been included in the folder for this purpose.

If we can assist you in any way to provide these materials, or if you have any questions about the study, please do not hesitate to call Dr. Marcia Turner, Associate Project Director, Transcript Study at (312) 702-8174 (collect) or Shirley Knight, Project Director, Transcript Study at (312) 702-8950 (collect).

Sincerely,

Barbara K. Campbell

Barbara K. Campbell
High School and Beyond
Project Director

Shirley Knight

Shirley Knight
Transcript Study
Project Director

AMERICAN ASSOCIATION of COLLEGIATE REGISTRARS and ADMISSIONS OFFICER
One Dupont Circle, N.W., Suite 330 • Washington, D. C. 20036 • (202) 293-91

June 15, 1987

Dear Colleague:

I would like to ask you to take a few moments from your busy schedule to provide the materials requested for a study that is being sponsored by the Center for Education Statistics (CES) and conducted by NORC, a social science research center at the University of Chicago.

The purpose of the Postsecondary Education Transcript Study is to obtain reliable and objective data concerning the types and patterns of courses taken by a nationally representative sample of students. This study is only one component of the National Longitudinal Studies Program conducted by CES since 1972.

The transcript study will collect transcript data for about 6,000 students who have attended approximately 2,100 postsecondary schools. Of course, the confidentiality of all data and the privacy of individuals and schools will be maintained according to the highest standards.

The information obtained in this study will make a valuable contribution to educational policy research on the relationship of postsecondary studies to occupational choice and success. Your cooperation and assistance in providing the transcripts in a timely manner to NORC will be greatly appreciated.

Sincerely,

John F. Collins, Jr.

John F. Collins, Jr.
President

rlh

NORC
Center For Education Statistics
National Longitudinal Studies Program
High School and Beyond

CES's Longitudinal Studies Program

The mandate of the Center for Education Statistics (CES) of the U.S. Department of Education includes the responsibility to "collect and disseminate statistics and other data related to education in the United States" and to "conduct and publish reports on specific analyses of the meaning and significance of such statistics" (Education Amendments of 1974 - Public Law 93-380, Title V, Section 501, amending Part A of the General Education Provisions Act).

Consistent with this mandate and in response to the need for policy-relevant, time-series data on a nationally representative sample of high school students, CES instituted the National Longitudinal Studies (NLS) program, a continuing long-term project. The general aim of the NLS program is to study longitudinally the educational, vocational, and personal development of high school students and the personal, familial, social, institutional, and cultural factors that may affect that development.

The NLS program was planned to make use of time-series databases in two ways: (1) each cohort is surveyed at regular intervals over a span of years, and (2) comparable data is obtained from successive cohorts, permitting studies of trends relevant to educational and career development and societal roles. The NLS program consists of two major studies: The National Longitudinal Study of the High School Class of 1972 (NLS-72) and High School and Beyond (HS&B).

High School and Beyond

High School and Beyond (HS&B) is a longitudinal study of the critical transition years as high school students leave the secondary school system to begin postsecondary education, work, and family formation. Its purpose is to provide information on the characteristics, achievements, and plans of high school students, their progress through high school, and the transition they make from high school to adult roles. Because of the breadth of the survey's coverage, data can be used to examine such policy issues as school effects, bilingual education, dropouts, vocational education, academic growth, access to postsecondary education, student financial aid, and life goals. High School and Beyond was designed to collect data that would be comparable to that of the National Longitudinal Study of the High School Class of 1972 (NLS-72).

In 1980, a national sample of over 30,000 sophomores and 28,000 seniors enrolled in 1,015 public and private schools participated in the Base Year Survey. During this stage of the study, students completed a cognitive test and a questionnaire about their high school experiences and plans for the future. In order to find out how plans have worked out or

changed, subsamples of the base-year students were asked to complete follow-up questionnaires in 1982, 1984 and 1986. The 1980 sophomore class also completed a cognitive test in 1982 when they were seniors. In addition, base-year data were compiled from such sources as school administrators, teachers, students' administrative records (transcripts), and parents of selected students.

In the spring of 1984 a consortium of university research centers sponsored a study of principals; guidance, vocational, and community service program counselors; and up to 30 teachers in each one of a sample of approximately 500 HS&B schools. Results of this survey, funded by the National Institute of Education, have become part of the HS&B database and permit researchers to describe the impact of the school environment on the educational process.

Postsecondary transcripts were collected for the senior cohort of HS&B in 1984. They contain reliable and objective information about the types and patterns of courses taken by students in colleges, graduate schools, and non-collegiate postsecondary institutions. The information has been merged with the expanding HS&B database. It will be possible for researchers to relate course-taking patterns to student characteristics available in the student questionnaire data files and to subsequent occupational choice and success.

A Financial Aid Records Study was conducted in 1985 for the senior cohort. Postsecondary schools attended by HS&B students provided data on the students' costs of attendance, student and family contributions, and financial aid packages. Guaranteed Student Loan records and Pell Grant information were collected from central data bases maintained in the Office of Education. Data from the three sources were then merged to provide a comprehensive profile of financial assistance.

Currently, records are being requested of Guaranteed Student Loans and Pell Grants that HS&B sophomores may have obtained. This financial aid information will be available to complement the postsecondary education transcripts. Hence, for the 1980 sophomore class, the Department of Education will have a complete record of high school experiences and past high school activities, including postsecondary schooling and financing.

A survey of the 1980 sophomore cohort's postsecondary transcripts is also underway. Some 2,100 postsecondary institutions are being asked to participate in this study. Like that of the senior cohort, the sophomore transcript study will provide information concerning the types and patterns of courses taken by students and will allow researchers to relate course-taking patterns to student characteristics available in the student questionnaire data files, and to subsequent occupational choice and success.

NATIONAL LONGITUDINAL STUDIES PROGRAM

High School and Beyond
A National Longitudinal Study for the 1980's

INSTRUCTIONS

Participation in the Postsecondary Education Transcript Study involves obtaining transcripts and related materials from your files and sending them to NORC, a social science research center at the University of Chicago. The steps on the following pages provide details on:

-- whose transcripts are requested

-- which school publications are requested

-- how to return materials to NORC

-- how to be reimbursed by NORC

Step 1: Review student checklist

The student checklist provides the names, in alphabetical order, of the students for whom copies of the transcript are being requested. In addition, other names (e.g., maiden, family, alternate spelling, etc.), social security numbers, and birthdates are provided as additional identifying information for many students. Please enter a check if you are enclosing a transcript(s) for a student. If you are unable to provide some or any of records for a student, please enter the reason in the space provided.

EXAMPLES:

"Never attended this school"

"Transcripts cannot be located at this time"

"Did not attend long enough to earn credit"

Two copies of the student checklist have been enclosed. Please return one copy with your checkmarks and any comments with the transcripts. The other copy is for your school's records.

Step 2: Retrieve and prepare transcripts

Locate and prepare (e.g., photocopy, generate a computer printout, etc.) a copy of each transcript for each student on the checklist.

Step 3: Label the transcripts

Affix the enclosed student labels to the back of the appropriate transcripts.

Step 4: Insert disclosure notices in each student's record file

Disclosure notices indicating the purpose for which student records were accessed for the transcript study are enclosed for your convenience.

Step 5: Obtain course catalog(s) or course list(s)

Obtain course catalog(s) or course list(s) describing the courses offered by your institution. Catalogs should be included for all programs and schools for which the student has been enrolled (e.g., the liberal arts college AND the law school). Please indicate on the checklist whether the current catalog(s) or course list(s) has been included in the package for return to NORC.

Step 6: Obtain grading system description

Obtain a copy of your school's official description of its grading system and/or other method of evaluating student performance. This might include, for example, an explanation of the meaning of letter grades (e.g., A,B...F), non-letter grading (e.g., Pass, High-Pass, Honors, etc.) , and/or other standard codes for the evaluation of student performance. In many instances, this would entail translation of grade designations to verbal (e.g., an "A" = ("Outstanding work"), or quantitative (e.g., "A"= "95—100") definitions.

Step 7: For reimbursement of expenses

If you would like to be reimbursed for the photocopying required for the transcripts or for other related expenses, please complete and return all copies of the enclosed voucher with the transcripts. One copy of the voucher will be returned with the check that will be issued upon receipt of the transcript package. If you have any questions regarding reimbursement, please call Dr. Marcia Turner, Transcript Study Associate Director (collect) at (312) 702-8174 or Shirley Knight, Transcript Study Project Director (collect) at (312) 702-8950.

Step 8: Assemble and send transcripts to NORC

A pre-paid, self-addressed envelope is enclosed for returning the transcripts and other related materials.

Please return all transcript study materials by August 1. If you encounter problems of any kind in regard to our request for transcripts, or you are unable to mail them by August 1 or shortly thereafter, please call Marcia Turner (collect) at (312) 702-8174 or Shirley Knight (collect) at (312) 702-8950.

ID: 002838 INST: SUNY AT STONY BK MAIN CAM

POSTSECONDARY EDUCATION TRANSCRIPT STUDY
EDUCATIONAL LONGITUDINAL STUDIES PROGRAM:
High School and Beyond

NORC 4414

STUDENT CHECKLIST: TRANSCRIPTS REQUESTED FOR 1 STUDENTS

PLEASE CHECK BOX IF COURSE CATALOGS OR LIST, AND A DESCRIPTION OF THE SCHOOL'S GRADING SYSTEM ARE ENCLOSED		IF NOT ENCLOSED, ENTER REASON (e.g. not available, non-existent, etc.)
Course catalog or current course list		
Description of grading system		

INSTRUCTIONS:

Please send transcripts or equivalent forms used for student program/performance for the students listed below.
See instruction folder for step-by-step details.
If you are unable to provide any student's transcript, please state the reason for each such student in the space provided.
Return this checklist with transcripts after having affixed the corresponding label on each.
Retain the second copy of the checklist for your records.

These students reported attending your school between 1982 - 1986

NORC USE ONLY			STUDENT'S NAME	ALTERNATE NAME (e.g. maiden, family, alternate spelling))	SOCIAL SECURITY NUMBER	BIRTHDATE	TRANSCRIPT ENCLOSED: enter X in box. NO TRANS ENCLOSED: enter reason (e.g., "never attended").
NUM	CASE ID	DISP					
001	1-999999-7		Lewis, Herman		111-22-1230	07/08/62	
002							
003							
004							
005							
006							
007							
008							

78

79

IMPORTANT: PLEASE CHECK REVERSE SIDE FOR ADDITIONAL NAMES.

Appendix B: Course Subject Codes in Numerical Order

CODING SYSTEM FOR COURSE AND PROGRAM OF STUDY CODING
FOR HS&B SOPHOMORE TRANSCRIPT SURVEY

PROGRAM/ COURSE CODE	CIP CODE	TITLE
01	01XXXX	AGRIBUSINESS & AGRICULTURAL PRODUCTION
02	02XXXX	AGRICULTURAL SCIENCES
03	03XXXX	RENEWABLE NATURAL RESOURCES
04	04XXXX	ARCHITECTURE & ENVIRONMENTAL DESIGN
05	05XXXX	AREA & ETHNIC STUDIES
06	06XXXX	BUSINESS & MANAGEMENT
07	0602XX	ACCOUNTING
08	0603XX	BANKING & FINANCE
09	07XXXX	BUSINESS & OFFICE
10	0706XX	SECRETARIAL & RELATED PROGRAMS (Note--this category does not include typing and general office, which are in 09 above)
11	08XXXX	MARKETING & DISTRIBUTION
12	09XXXX	COMMUNICATIONS
13	0904XX	JOURNALISM
14	10XXXX	COMMUNICATIONS TECHNOLOGIES
15	11XXXX	COMPUTER & INFORMATION SCIENCES
16	1102XX	COMPUTER PROGRAMMING
17	1103XX	DATA PROCESSING
18	12XXXX	CONSUMER, PERSONAL & MISCELLANEOUS SERVICES
19	13XXXX	EDUCATION
20	131201	ADULT & CONTINUING EDUCATION
21	131202	ELEMENTARY EDUCATION
22	131203	JUNIOR HIGH EDUCATION
23	131204	PRE-ELEMENTARY EDUCATION
24	131205	SECONDARY EDUCATION
25	14XXXX	ENGINEERING
26	1408XX	CIVIL ENGINEERING
27	141001	ELECTRICAL, ELECTRONICS & COMMUNICATIONS ENGINEERING
28	1419XX	MECHANICAL ENGINEERING
29	15XXXX	ENGINEERING & ENGINEERING RELATED TECHNOLOGIES
30	16XXXX	FOREIGN LANGUAGES
31	160501	GERMAN
32	160901	FRENCH
33	160905	SPANISH
34	17XXXX	ALLIED HEALTH
35	170605	PRACTICAL NURSING
36	18XXXX	HEALTH SCIENCES
37	1811XX	NURSING
38	19XXXX	HOME ECONOMICS
39	20XXXX	VOCATIONAL HOME ECONOMICS
40	220101	LAW
41	23XXXX	LETTERS
42	230401	COMPOSITION
43	230701	AMERICAN LITERATURE
44	230801	ENGLISH LITERATURE

45	25XXXX	LIBRARY & ARCHIVAL SCIENCES
46	26XXXX	LIFE SCIENCES
47	27XXXX	MATHEMATICS
48	279999	CALCULUS
49	28XXXX	MILITARY SCIENCES (includes 29XXXX--Military Technologies)
50	31XXXX	PARKS & RECREATION
51	32XXXX	FUNCTIONAL SKILLS (includes 32XXXX - 37XXXX: Basic Skills, Citizenship/Civic Activities, Health-Related Activities, Interpersonal Skills, Leisure and Recreational Activities, Personal Awareness)
52	38XXXX	PHILOSOPHY & RELIGION
53	39XXXX	THEOLOGY
54	40XXXX	PHYSICAL SCIENCES
55	4005XX	CHEMISTRY
56	400601	GEOLOGY
57	4008XX	PHYSICS
58	41XXXX	SCIENCE TECHNOLOGIES
59	42XXXX	PSYCHOLOGY
60	43XXXX	PROTECTIVE SERVICES
61	44XXXX	PUBLIC AFFAIRS
62	4407XX	SOCIAL WORK (includes Medical Social Work)
63	45XXXX	SOCIAL SCIENCES
64	4502XX	ANTHROPOLOGY
65	4506XX	ECONOMICS
66	4507XX	GEOGRAPHY
67	4508XX	HISTORY
68	4510XX	POLITICAL SCIENCE & GOVERNMENT
69	4511XX	SOCIOLOGY
70	46XXXX	CONSTRUCTION TRADES
71	47XXXX	MECHANICS & REPAIRERS
72	48XXXX	PRECISION PRODUCTION includes 21XXXX--Industrial Arts)
73	49XXXX	TRANSPORTATION & MATERIAL MOVING
74	50XXXX	VISUAL & PERFORMING ARTS
75	5003XX	DANCE
76	5007XX	FINE ARTS
77	5009XX	MUSIC
78	24XXXX	LIBERAL/GENERAL STUDIES (includes 30XXXX--Multi/ Interdiscipline studies)
95	999995	UNCODEABLE
96	XXXXXX	TRANSFER COURSES
99	XXXXXX	MISSING