DOCUMENT RESUME

ED 411 325 TM 027 708

AUTHOR	Legum, Stanley; Caldwell, Nancy; Davis, Bryan; Haynes, Jacqueline; Hill, Telford J.; Litavecz, Stephen; Rizzo, Lou; Rust, Keith; Vo, Ngoan
TITLE	The 1994 High School Transcript Study Technical Report.
INSTITUTION	Westat, Inc., Rockville, MD.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-97-262
ISBN	ISBN-0-16-049170-3
PUB DATE	1997-08-00
NOTE	203p.; For the High School Transcript Study tabulations, see TM 027 709.
AVAILABLE FROM	U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328.
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Evaluative (142)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*Academic Achievement; Academic Records; *Core Curriculum; *Course Selection (Students); Data Collection; Data Processing; Educational Policy; Educational Trends; *High School Students; High Schools; National Surveys; Policy Formation; Research Design; *Research Methodology; Sample Size; *Sampling; Tables (Data); Trend Analysis
IDENTIFIERS	*High School Transcript Study; National Assessment of Educational Progress; National Center for Education Statistics

ABSTRACT

The 1994 High School Transcript Study (HSTS) provides the Department of Education and other policymakers with information about current course offerings and students' course-taking patterns in the nation's secondary schools. One objective was to determine changes in course offering and selection patterns since the previous studies in 1982, 1987, and 1990. Another research objective was to compare course-taking patterns to results from the National Assessment of Educational Progress (NAEP), an assessment of educational achievement nationwide. In 1994, transcripts were collected for more than 25,000 students who graduated from high school that year. These students were from 340 schools that participated in the NAEP. This technical report documents the procedures used to collect and summarize the data. An accompanying volume provides tables of findings, and another contains the data file user's manual. The following sections are included: (1) "Executive Summary"; (2) "Background: Sample Design"; (3) "Selection of Schools and Students for the 1994 High School Transcript Study"; (4) "Data Collection Procedures"; (5) "Data Processing Procedures"; (6) "Weighting and Estimation of Sampling Variance"; (7) "1994 High School Transcript Study Data Files"; and (8) "References." Five appendixes provide supplementary information about the study methodology. (Contains 30 tables, 2 figures, 23 exhibits, and 18 references.) (SLD)

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report August 1997

The 1994 High School Transcript Study Technical Report

U.S. Department of Education
Office of Educational Research and Improvement NCES 97-262

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report

August 1997

The 1994 High School Transcript Study Technical Report

Stanley Legum
Nancy Caldwell
Bryan Davis
Jacqueline Haynes
Telford J. Hill
Stephen Litavecz
Lou Rizzo
Keith Rust
Ngoan Vo
Westat, Inc.

Steven Gorman, Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement

NCES 97-262

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-048170-3

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

Office of Educational Research and Improvement
Ramon C. Cortines
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208–5574

August 1997

The NCES World Wide Web Home Page address is
http://www.ed.gov/NCES/

Suggested Citation
U.S. Department of Education. National Center for Education Statistics. *The 1994 High School Transcript Study Technical Report,* NCES 97-262, by Stanley Legum, Nancy Caldwell, Bryan Davis, Jacqueline Haynes, Telford J. Hill, Stephen Litavecz, Lou Rizzo, Keith Rust, and Ngoan Vo. Project Officer, Steven Gorman. Washington, DC: 1997.

Contact:
Steven Gorman
(202) 219–1937

THE 1994 HIGH SCHOOL TRANSCRIPT STUDY
TECHNICAL REPORT

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

List of Tables

TABLE OF CONTENTS (CONTINUED)

List of Tables (continued)

List of Figures

List of Exhibits

TABLE OF CONTENTS (CONTINUED)

List of Exhibits (continued)

1. EXECUTIVE SUMMARY

The 1994 High School Transcript Study (HSTS) was conducted by Westat, Inc. for the U.S. Department of Education's National Center for Education Statistics. This study provides the Department of Education and other educational policymakers with information regarding current course offerings and students' course-taking patterns in the nation's secondary schools. Since similar studies were conducted of course-taking patterns of 1982, 1987, and 1990 graduates, one research objective was to study changes in these patterns. In particular, the data from the 1994 study permit analysts to investigate the impact of the Core Curriculum recommended by the National Commission on Excellence in Education in 1983.[1] Another research objective was to compare course-taking patterns to study results on the 1994 National Assessment of Educational Progress (NAEP). NAEP is a Federally-funded, ongoing, periodic assessment of educational achievement in the various subject areas and disciplines taught in the nation's schools. Since 1969, NAEP has gathered information about the levels of educational achievement of 9-, 13-, and 17-year-old students across the country.

In the Summer and Fall of 1994, Westat collected high school transcripts from over 25,000 students who graduated from American high schools in 1994. These students attended 340 schools that were sampled for NAEP in 1994. The sample of schools was nationally representative of all schools in the United States, and the sample of students was representative of graduating seniors from each school. While the NAEP sample included students who were enrolled in the 12th grade at the time of the NAEP sampling, the transcript study included only those students whose transcripts indicated that they graduated between January 1, 1994 and November 21, 1994, the date the final transcripts were collected.[2]

Approximately 90 percent of the students included in the transcript study also participated in NAEP assessments in 1994. The remaining students were sampled specifically for the transcript study either because their schools did not agree to participate in the NAEP study, or because the schools

---

[1] In its report to the Secretary of Education entitled "A Nation at Risk," the National Commission on Excellence in Education's first recommendation was "We recommend *that State and local high school graduation requirements be strengthened and that,* at a minimum, all *students seeking a diploma be required to lay the foundations in the Five New Basics by taking the following curriculum during their 4 years of high school: (a) 4 years of English; (b) 3 years of mathematics; (c) 3 years of science; (d) 3 years of social studies; and (e) one-half year of computer science. For the college-bound, 2 years of foreign language in high school are strongly recommended in addition to those taken earlier."* For the sake of brevity, this recommended set of courses is referred to as "the Core Curriculum."

[2] An analysis of the 1990 High School Transcript Study data showed that only 0.17 percent of the students with known graduation dates graduated between September 1 and December 31 and that only 1.13 percent graduated in July and August. Approximately 90 percent of the transcripts were collected in August and September 1994 and the remainder in October and November.

participated in NAEP but did not retain their administration materials linking student identification numbers to student names.

The 1994 High School Transcript Study is documented in three reports:

- The 1994 High School Transcript Study Technical Report - This is the document you are now reading. It documents the procedures used to collect and summarize the data.
- The 1994 High School Transcript Study Tabulations - The Tabulations volume provides copious tables summarizing the course-taking patterns of 1994 high school graduates and comparing them to those of their counterparts in 1982, 1987, and 1990. It also provides tables describing the relationship of the course taking patterns of 1994 graduates to their proficiencies in reading, geography, and history as measured by the 1994 National Assessment of Educational Progress.
- The 1994 High School Transcript Study Data File User's Manual - The Data File User's Manual provides a detailed description of all publicly released data files that were produced by the study.

The Coding System

In order to compare transcripts from different schools, it is necessary to code each of the courses entered on the transcripts using a common course coding system. The coding system employed for this purpose was a modification of the *Classification of Secondary School Courses* (CSSC) (Ludwig, et al.). The CSSC, which contains approximately 2,000 course codes, is a modification of the Classification of Instructional Programs (CIP) that is used for classifying college courses (Morgan, et al.). Both systems use a three-level, six-digit system for classifying courses. The CSSC uses the same first two levels as the CIP, which are represented by the first four digits of each code.[3] The third level of the CSSC (the fifth and sixth digits of the course code) unique to the CSSC and represents specific high school courses.

The CSSC also uses an additional one-digit "disability" flag and a one-digit "sequence" flag. The first flag indicates whether a course is open to all students or is restricted to disabled students. The sequence flag indicates whether a course is part of a sequence of courses and, if so, its place in that

[3] Actually, the CSSC uses the first two levels of the CIP as it existed in 1982. The CIP has undergone some modification since then. In addition, three sets of codes at the top level have been added to the CSSC to provide a means of classifying courses specifically designed for disabled students.

sequence. The disability flag was added to the CSSC during the 1987 transcript study. The sequence flag was added during the 1990 study.

During the 1987, 1990, and 1994 studies, courses appearing on student transcripts were coded to indicate whether they were transfer courses, offered off campus, honors or above grade-level courses, remedial or below grade-level courses, or designed for students with limited English proficiency (LEP) and/or taught in a language other than English.

We used course catalogs and related materials and information from the participating schools to determine the codes assigned to each course. We also entered the grades and credits received for each course and standardized them into a consistent system.

Student Information

Information was gathered for all students included sex, grade level, birth year, birth month, graduation status, race/ethnicity, and whether or not the student had an Individualized Education Program (IEP) or a Limited English Proficiency (LEP) or received Chapter 1 services. When it was available, we also obtained the date of entry to the school, the graduation date, type of diploma, number of days absent in each of 4 years (9th grade, 10th grade, 11th grade, and 12th grade), grade point average, and class rank. In addition, we listed all awards and standardized tests taken by each student as reflected on the transcript.

In some cases, more than the basic information was obtained. The following additional information, as reported by school personnel, was collected for disabled students: grade-level equivalent performance in English and mathematics, proportion of placement in mainstream classes, type and severity of disability, and type of special services provided.

Students with limited proficiency in English were also included in the study. The following additional information, as reported by school personnel, was collected for students with limited English proficiency: English and mathematics grade levels, percentage of the day spent in special language programs, native language, information on the student's linguistic environment, type of specialized instruction, number of years that the student was in a special language program, and the student's ability to speak, understand, read, and write English.

Student transcript data were weighted for the purpose of making estimates of course-taking patterns by students in the class of 1994 nationwide. Five sets of weights were created:

- Weights for all students who participated in the transcript study; i.e., for whom a transcript was received and coded;
- Four sets of "linked" weights for students who participated in both the transcript study and NAEP. Since students participating in NAEP were selected to participate in the assessment of a particular subject, separate weights were developed for the students in each subject-specific assessment:
  - Weights for students who participated in the transcript study and the NAEP reading assessment;
  - Weights for students who participated in the transcript study and the NAEP geography assessment;
  - Weights for students who participated in the transcript study and the NAEP history assessment; and
  - Weights for students who participated in the transcript study but were excluded from NAEP because of a disability or limited English proficiency.

In each set of weights, the final weight attached to an individual student record reflected two major aspects of the sample design and the population surveyed. The first component, the base weight, reflects the probability of selection in the sample (the product of the probability of selecting the primary sampling unit (geographic area), the probability of selecting the school within the primary sampling unit, and the probability of selecting the student within the school). The second component resulted from the adjustment of the base weight to account for nonresponse within the sample and to ensure that the resulting survey estimates of certain characteristics (race/ethnicity, size of community, and region) conformed to those known reliably from external sources.

Estimation of sampling errors was performed by an application of the jackknife replication procedure.[4] A set of 62 replicate weights was attached to each record, one for each replicate. Variance estimation was performed by repeating the estimate procedure 63 times, once using the original full set of sample weights, and once each for the set of 62 replicate weights. The variability among replicate estimates was used to derive an approximately unbiased estimate of the sampling variance. This procedure

[4] Because the HSTS used a multistage sampling design and because estimates were adjusted by both poststratification and weighting adjustments, observations on different students are not independent. For this reason, variance estimation formulas which assume independence will underestimate the sample variability. As discussed in Chapter 6, jackknife replication provides reliable variance estimates for data like those in the HSTS.

was used to obtain sampling errors for a large number of variables for the whole population and for specified subgroups.

In general, the variability was small compared to the size of the estimates, although this is not true in cases of infrequently taken courses in the smaller subpopulations. For example, the percentage of white students taking geometry is estimated at 72.38, with a standard error of 1.56 (a ratio of 0.02), while the percentage of Native Americans taking calculus is estimated at 3.75, with a standard error of 1.23 (a ration of 0.33). These and thousands of other estimates are presented in *The 1994 High School Transcript Study Tabulations* (NCES 97-260).

Data Files

The study has produced a set of eight data files that are available on a restricted use basis:

- The Master CSSC File – The Classification of Secondary School Courses (CSSC) including all modifications made to the original (1982) CSSC during the 1987, 1990, and 1994 transcript studies. This file has separate variables for the CSSC code, the disability flag, the sequence flag, and the course title.
- The Course Offerings File -- Provides a comprehensive listing of the courses offered in the 340 schools included in the study. A code from the CSSC has been associated with each course title.
- The School File – Provides detailed information on the schools from which the students were sampled.
- The Student File – Provides demographic information on all students included in the study, as well as weighting data and summaries of their course-taking histories.
- The Linked Weights File -- Provides weights for use when performing analyses relating transcript data to NAEP assessment results.
- The IEP/LEP Questionnaire File – Provides information on the disabled students and students with limited English proficiency who are included in the study.
- The Test and Honors File -- Provides a list of honors and standardized test results that were included on the transcripts.
- The Transcript File – Provides a complete list of all courses appearing on the transcripts of students included in the study.

Three additional NAEP assessment files contain proficiency scores for each student who completed NAEP. These are:

- The 1994 NAEP Reading Data File;
- The 1994 NAEP Geography Data File; and
- The 1994 NAEP History Data File.

These three files contain NAEP scores for 1994 graduates who participated in both the specific NAEP assessment and the transcript study.

This report describes the 1994 NAEP sample (Chapter 2) in so far as it relates to the High School Transcript Study. It then describes the school and student sampling issues that are specific to the transcript study (Chapter 3). Chapter 4 provides a detailed description of the data collection procedures. Chapter 5 describes the data entry and course coding operations. Chapter 6 provides a full description of how we weighted the data so that they can be used to predict national totals. This description documents the need for separate sets of weights for analysis of transcript data alone and for the joint analysis of transcript and NAEP data, as well as the techniques we used to produce each set of weights. Chapter 7 provides a short summary of each of the data files produced by the study. A list of references appears as Chapter 8.

There are also several appendixes at the end of the report that give examples of forms used in the schools, the study questionnaires, and the 1994 additions to the *Classification of Secondary School Courses*.

2. BACKGROUND: SAMPLE DESIGN

The 1994 High School Transcript Study (HSTS) was designed to allow an analysis of the course-taking patterns of students who graduated from American high schools in 1994. In addition, it was designed so that data on students' course-taking patterns can be linked to the 1994 National Assessment of Educational Progress (NAEP) assessment results. As noted earlier, NAEP is a Federally-funded, ongoing, periodic assessment of educational achievement in the various subject areas and disciplines taught in the nation's schools. Since 1969, NAEP has gathered information about levels of educational achievement of 9-, 13-, and 17-year-olds across the country. Since studies similar to the 1994 HSTS were conducted on 1982, 1987, and 1990 graduates, changes in these patterns and relationships to NAEP performance in these years can be studied.[5]

The HSTS used a subsample of schools from the 1994 NAEP assessment for grade 12/age 17 students. Although HSTS used the NAEP target sample of students in these subsampled schools, the HSTS sample was restricted to 12th graders, while NAEP assessed both 12th graders and students who were 17 years old (students born in 1976). This chapter describes aspects of the 1994 NAEP sample design that affect the transcript study. Chapter 3 describes aspects of the selection of schools and students that are specific to the transcript study.

2.1 1994 NAEP Sample Design

The samples for the 1994 NAEP assessment were selected using a complex, multistage sample design that involved sampling students from selected schools within 94 selected geographic areas, called primary sampling units (PSUs), across the United States.

The sample design had four stages of selection:

(1) Selection of geographic PSUs (counties or groups of counties);
(2) Selection of schools within PSUs;

[5] The 1987 and 1990 transcript data were collected by Westat in coordination with the 1987 and 1990 NAEP (Thorne et al., 1989; Legum, et al., 1993). The 1982 data were collected by the National Opinion Research Center as part of the High School and Beyond project (Jones, et al., 1983a).

(3) Assignment of session types to schools;[6] and
(4) Selection of students for session types within schools.

The main NAEP sample represented all grade 12 students in the United States. Within the main sample, private schools and public schools with moderate or high enrollment of black or Hispanic students were oversampled to increase the reliability of estimates for students in private schools and in these two minority groups.

2.2 Selection of Primary Sampling Units (PSUs)

In the first stage of sampling, the United States (the 50 states and the District of Columbia) was divided into geographic primary sampling units. Each PSU met a minimum size requirement (a 1990 census population of at least 60,000 in the Northeast and Southeast and 45,000 in the Central or West regions) and comprised a metropolitan statistical area (MSA), a single county, or (more often in the case of nonMSA PSUs) a group of contiguous counties. In the case of New England MSAs, which are not formed from whole counties, the corresponding New England County Metropolitan Areas, which are defined in terms of whole counties, were designated as the PSUs. Each PSU was contained entirely within one of the four geographic regions defined in Table 2-1. Each region contains about one-fourth of the U.S. population. These regions were used to stratify the sample of PSUs, ensuring that each region was adequately represented in the various assessment samples.

In a few cases, a metropolitan statistical area crossed region boundaries. Such MSAs were split into two or more PSUs as necessary. For example, the Cincinnati OH-KY-IN MSA was split into the Cincinnati OH-IN PSU in the Central region and the Cincinnati KY PSU in the Southeast region.

[6] There were two distinct types of session: Reading and Geography/History. Different students participated in each session. A student in the Geography/History session received either a Geography assessment or a History assessment, but not both.

Table 2-1. Geographic regions used for stratification

Northeast	Southeast	Central	West
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia*	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia*			Utah
			Washington
			Wyoming

*That part of Virginia that is part of the Washington, DC-MD-VA metropolitan area is included in the Northeast region; the remainder of the state is included in the Southeast.

The 22 largest PSUs in the United States were included with certainty (that is, with probability = 1). The remaining smaller PSUs were not guaranteed to be selected for the sample (that is, they were included with probability < 1). These were grouped into a number of noncertainty strata and one PSU was selected from each stratum. Within each major stratum or subuniverse, further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics, yielding 72 strata.

The strata were defined so that the aggregate of the measures of size of the PSUs in a stratum was approximately equal for each stratum. The size measure used was the population from the 1990 Census. The characteristics used to define strata were the percentage minority population, percentage change in total population since 1980, per capita income, percentage of persons age 25 or over with college degrees, percentage of persons age 25 or over who completed high school, and the civilian unemployment rate. Up to four of these characteristics were used in one subuniverse. For each subuniverse, the characteristics used were chosen by modeling PSU-level mean reading proficiency scores for 1988, 1990, and 1992. Then one PSU was selected with probability proportional to size from each of the 72 noncertainty strata. That is, within each stratum, a PSU's probability of being selected was proportional to its population.

The final sample of 94 PSUs was drawn from a population of about 1,000 PSUs. Primarily because of the use of MSAs as PSUs (they varied greatly in size), PSUs varied considerably as to their probability of selection. In each region, noncertainty PSUs were classified as metropolitan (MSA) or nonmetropolitan (nonMSA). The 36 selected noncertainty MSA PSUs had probabilities ranging from 0.023 to 0.580, while the 36 nonMSA PSUs had probabilities ranging from 0.029 to 0.108. Parts of 41 states were included in the main sample PSUs. Ninety-four PSUs were selected for the main NAEP sample (22 certainty and 72 noncertainty). These same PSUs were used for the HSTS sample. The major strata, or subuniverses of noncertainty PSUs, are shown in Table 2-2.

Table 2-2. Noncertainty PSU strata

Region	Number of strata for MSA PSUs	Number of strata for nonMSA PSUs	Total
Northeast	6	4	10
Southeast	12	12	24
Central	8	12	20
West	10	8	18
Total	36	36	72

2.3 Selection of NAEP Schools

After the PSUs were selected, the next step was to select the schools within the PSUs. For the second stage of sampling, a frame list was prepared of all schools with at least one of the four grades 9 through 12. This list included all public schools (including Bureau of Indian Affairs and Department of Defense schools) and private schools with these grades in the 94 sampled PSUs. There were 5,178 public and 5,406 private schools on the final second stage sampling frame.

The lists of schools were obtained from several sources. Information on regular public, Bureau of Indian Affairs, Department of Defense, Catholic, and other private schools was obtained from the 1992 list of schools maintained by Quality Education Data, Inc. (QED).[7]

[7] Quality Education Data, Inc. (Denver, CO) (QED) is a privately maintained database of public and private schools in the United States that provides an annual listing of all schools and school districts in the United States, released in November of each year. This listing

Supplementary lists of private schools were obtained from three sources and added to the QED list of private schools. This supplementation was undertaken because previous studies have revealed that the QED list is somewhat deficient in its coverage of non-Catholic private schools.

The first supplementary private school listing source used was the Private School Survey (PSS) developed for the National Center for Education Statistics' 1988 School and Staffing Survey. This list was restricted to a sample of counties selected for the survey. Certain of these counties, generally large in population, were also included, independently by chance, in the NAEP sample PSUs. The schools from such counties were added to the NAEP frame after steps were taken to eliminate duplicates with the QED list of private schools. The second and third sources were lists generated clerically from the yellow pages of telephone directories from metropolitan areas included in the 1992 and 1994 NAEP PSU samples. These lists were matched against each other and against other private school sources to eliminate duplicates. The supplementary lists contributed 2,896 of the 5,406 private schools on the sampling frame.

Each public school that was considered high minority (i.e., with over 15 percent black and/or Hispanic enrollment) was given double the probability of selection of a public school, not considered high minority, of similar size in the same PSU. Such high-minority schools were oversampled to enlarge the sample of black and Hispanic students, thereby enhancing the reliability of estimates for these groups. For a given overall size of sample, this procedure reduces slightly the reliability of estimates for all students as a whole and for those not black or Hispanic.

In NAEP, each private school was given triple the probability of selection of a low-minority public school of similar size from the same PSU. These greater probabilities of selection were used to ensure adequate samples of private school students in order to allow the derivation of reliable NAEP estimates for such students. In HSTS, however, the oversampling of private schools was reversed by taking a private school subsample from the NAEP sampled schools at only one-third the sampling rate of the corresponding public school sample (see Chapter 3).

corresponds to the previous school year. It includes information about each school's name, mailing address, location address, district name, FIPS state number, Office of Education district number, number of students, number of teachers, grades served, and other sociodemographic data.

The QED files do not contain schools that opened between 1992 and the assessment dates. Therefore, special procedures were implemented to be sure that the NAEP assessment represented students in new public schools. Small school districts, which generally contained only one eligible school for a given age class, were treated differently from large school districts, which generally contained more than one eligible school for a given age class. In small school districts, the schools selected for a given age class were thought to contain all students in the district that were eligible for the assessment. Districts containing these schools were asked if other schools with the appropriate grades for the assessment existed, and if so, they were automatically included in the assessment. For large school districts, a district-level frame was constructed from the schools on the QED file that were eligible for one of the national assessments. Then districts were sampled systematically with probabilities proportional to a measure of size. In most cases, the measure of size was total district enrollment, but in very small districts a minimum measure of size was used. Each sampled district was asked to update lists of eligible schools according to information on the QED files. Frames of eligible new schools were then constructed separately for each age class, and separate samples of new schools were selected systematically with probability proportional to eligible enrollment using the same sampling rates as for the old schools. Four new schools were added to the age 17 main sample.

In a few PSUs where school refusals were relatively heavy for a particular sample, substitute school selections were made, replacing the refusals (to the extent feasible) with schools from within the same PSU and similar in size, affiliation (public, Catholic, or other private), grade span, and minority composition. Two substitute schools were included in the age 17 main sample.

2.4 Assignment of Sessions to Schools

There were two session types: reading and history/geography. The larger schools were assigned 6 sessions, 3 of which were reading and 3 history/geography. Smaller schools were assigned from 1 to 5 sessions, based on the number of eligible students. If 2 sessions were assigned, 1 was reading and 1 was history/geography. If 3 sessions were assigned, 1 or 2 were reading, with the remainder history/geography. If 4 sessions were assigned, 2 were of each session type. If 5 sessions were assigned, 2 were of one session type and 3 were of the other.

Schools with less than 20 eligible students were assigned only 1 session type. This single session was randomly assigned to be either a reading session or a history/geography session, with equal probability assigned to each outcome.

2.5 Sampling Students

In the fourth stage of sampling, a consolidated list was prepared for each school of all grade-eligible and age-eligible students of the age class for which the school was selected. A systematic selection of eligible students was made from this list (unless all students were to be assessed) to provide the target sample size. For schools assigned to more than a single session type (the vast majority), students were assigned by Westat district supervisors to one of the various session types in a systematic random manner.

A maximum sample size of 200 students was set for each school. In those schools that, according to information on the frame, had fewer than 200 eligible students, each eligible student enrolled at the school was selected in the sample for one of the sessions assigned to the school. In other schools, a sample of students was drawn, and then students were assigned to sessions as appropriate.

The sample of students to be selected in each school was derived in the following manner. On the basis of data obtained from the School Characteristics and Policies Questionnaire (or the sample frame, if the questionnaire data were not obtained in time) an estimate of the number of eligible students was established for each school. The estimated number of grade-plus-age-eligible students was used for this purpose (grade-eligible students were in 12th grade; age-eligible students were 17 years old in calendar year 1994). A Session Assignment Form was generated for each school, showing the line numbers of the students to be selected and indicating the type of session to be taken by each student. The line numbers were generated using a sampling interval designed to give the appropriate sample size for each school.[8] Thus, the overall sampling interval was 1.0 for schools in which all eligible students were to be assessed. The appropriate sampling interval was specified for schools with more than 200 eligible students.

[8] A line number was a sequential number assigned to a student in the order in which he or she appeared on the enrollment list provided by each school.

If the field worker found that the line numbers, when applied to the numbered list of eligible students assembled in the field for each school, generated a sample in excess of 240 students (120 percent of the maximum sample size), he or she called Westat's central office. New line numbers based on the actual number of eligible students were generated on a personal computer at the central office and relayed to the field worker. A similar revision to the line numbers was made in a school with a sampling interval in excess of 1.0 and eligible enrollment less than 80 percent of that initially estimated. In this case, the sample size was increased to the appropriate level. This procedure provided a suitable compromise between control over the sampling rate within each school and operational autonomy and flexibility for field workers.

Note that, in all cases, sampling intervals were generated in Westat's central office and stored for use in sample weighting. Field workers were not required to derive or record within-school sampling rates.

2.6 Students not Included in the Assessment

Once the sample of students was selected, school staff were asked to identify any students with an Individual Education Plan, for reasons other than being gifted and talented, and students classified as limited English proficient. A questionnaire, the IEP/LEP Student Questionnaire, was then distributed to the school staff member most knowledgeable about the student, as described in Section 4.5. The questionnaire collected information about the student's disability/language proficiency and any special services provided by the school.

School staff were also asked to determine whether any of the students identified as IEP or LEP could not participate meaningfully in the assessment. These students were not invited to the assessment and were coded as "excluded" to distinguish them from absent students. Transcripts for these students are, however, included in the transcript study.

3. SELECTION OF SCHOOLS AND STUDENTS FOR THE 1994 HIGH SCHOOL TRANSCRIPT STUDY

The purposes of the 1994 High School Transcript Study (HSTS) were to gather data on a nationally representative sample of students who graduated from American high schools in 1994 and to gather data that can be linked to NAEP results. For the HSTS school sample to be as representative as possible, it included a subsample of all schools with 12th grades that were selected for NAEP, regardless of whether they participated in NAEP. A representative sample of students was included from each school. When possible, the students selected for the transcript study were the same as those selected for NAEP. When this was not possible, a systematic sample of students was drawn from the school. The school sample and the student sample are described in detail in the following two sections.

3.1 School Sample

As discussed in Chapter 2, the 1994 NAEP sample included both schools with 12th grades and schools without 12th grades if 17-year-old students were enrolled. The 1994 HSTS sample, however, included only schools selected for the NAEP main sample that had 12th-grade classes. There were 538 eligible schools that satisfied this criterion, of which 379 were public and 159 were private. In the next step of selection, a subsample of 333 public schools was drawn from the list of eligible NAEP public schools (a sampling rate of 88.1 percent), and a subsample of 47 private schools was drawn from the list of eligible NAEP private schools (a sampling rate of 29.4 percent). Each subsample was an equal probability systematic sample from the list of eligible NAEP sample schools (in their original frame order). The private schools were sampled at a lower rate to offset the tripled probability of selection they received in the NAEP sample. (An oversample of private schools was considered important for the NAEP sample, but was not considered desirable for the HSTS sample. Because private schools tend to be smaller than public schools, the collection cost per transcript is higher in private schools than public schools.)

In order to maintain as many links as possible with NAEP scores, where schools refusing to participate in NAEP were replaced by substitute schools, the substitute schools, not the refusals,

were asked to participate in the HSTS. Of the 379 schools in the original sample, 340 participated in the HSTS survey.

3.2 Student Sample

For schools participating in both NAEP and HSTS, the same students were, where possible, included in the two samples. For privacy reasons, the only means of identifying the names of students participating in NAEP is a list left in the school office. Since the NAEP assessments were administered from January through April 1994, the schools were asked to retain the NAEP administration schedules until the HSTS data collection in the Summer and Fall of 1994.[9] Only three schools did not retain their NAEP administration schedule.[10]

For schools that participated in NAEP but were missing their administration schedules, and for schools that agreed to provide transcripts but did not participate in the NAEP assessment, the field workers sampled the students using the following rules:

- If 60 or fewer students were in the senior class, then transcripts were collected for the entire class.
- If more than 60 students were in the senior class, then the field worker drew a systematic random sample of 50 transcripts.

To draw a sample, the field worker obtained a complete list of students in the senior class, numbered each student sequentially, and then entered the number of students in the class and the number of transcripts needed onto a sampling form. After determining the number of students in the senior class, the field worker calculated a sampling interval. A random start was drawn from a list of random numbers, and a systematic sample was drawn based on the random start and the sampling interval. The field worker then wrote the names of the sampled students on a Transcript Request Form

[9] NAEP asked schools to retain the administration schedules until the end of the school year in case it became necessary to use them to resolve ID-related questions. For reasons of confidentiality, the schools that were not in the transcript study were requested to destroy these materials by June 30, 1994.

[10] This was a major improvement in the retention rate from previous transcript studies. In 1990, only 204 of 283 schools that participated in both schools retained the administration schedules. In 1987, only 192 of 363 schools participating in both studies retained the administration schedules. The reasons for the improved retention rate in the current study are (1) earlier notification of the schools to retain the administration schedules and (2) earlier collection of the transcripts.

(TRF) and gave it to the school staff to draw the transcripts. The TRF also provided a place to record the student's graduation status, sex, race, birth month, and birth year. The field worker removed the students' names before returning a copy of the TRF to Westat along with the transcripts. A copy of the TRF is included as Exhibit 4-9.

A total of 28,815 students were selected for inclusion in the HSTS. Of these, 25,904 students were from schools that maintained their NAEP administration schedules and are identified by their NAEP booklet numbers. Another 216 students were from schools that participated in NAEP but had lost the link between student names and NAEP booklet numbers, and 2,695 were from schools that did not participate in NAEP. A detailed description of sampling results and nonresponse rates is presented in Chapter 6.

Table 3-1 displays the number of eligible schools in the sample and the number and percentage of schools from which we collected transcripts by linking category.

Table 3-1. Response rates of eligible schools by linking category

School participation status	Number of schools in sample	Number of schools where data were collected	Percent of schools where data were collected
School participated in NAEP -- IDs linked to NAEP IDs	292	280	95.9
School participated in NAEP -- IDs not linked to NAEP IDs	3	3	100.0
School did not participate in NAEP	84	57	67.9
Total sampled schools	379	340	89.7

Table 3-2 displays the number of sampled students in the participating schools and the number and percentage of complete transcripts of graduates that were processed.

Table 3-2. Response rates of students in eligible participating schools

School participation status	Number of students in sample	Number of transcripts of graduates collected	Percent of transcripts collected
School participated in NAEP -- IDs linked to NAEP IDs	25,904	22,716	87.7
School participated in NAEP -- IDs not linked to NAEP IDs	216	174	80.6
School did not participate in NAEP	2,695	2,604	96.6
Total	28,815	25,494	88.5

Because sampling was performed in most schools using a list of seniors rather than a list of graduates, not all sampled students were in fact graduates. Only graduates, however, were eligible for the transcript study. We know that 25,581 sampled students actually graduated and that 2,717 did not. Of the remaining 517 students, we imputed 454 as graduates and 63 as not. Thus, we collected and processed 25,494 transcripts of graduates from a sample of 26,045. That is, we were able to obtain 97.9 percent of the transcripts of eligible students. Table 3-3 displays the response rates for graduates in the eligible participating schools.

Table 3-3. Response rates of graduates

School participation status	Known graduates	Imputed graduates	Known and imputed graduates	Transcripts of graduates collected	Percent of transcripts of known graduates collected	Percent of transcripts of known and imputed graduates collected
School participated in NAEP -- IDs linked to NAEP IDs	22,799	431	23,230	22,716	99.6	97.8
School participated in NAEP -- IDs not linked to NAEP IDs	174	28	202	174	100.0	86.1
School did not participate in NAEP	2,608	5	2,613	2,604	99.8	99.7
Total	25,581	464	26,045	25,494	99.7	97.9

4. DATA COLLECTION PROCEDURES

4.1 Training NAEP Field Supervisors as Data Collectors

The field workers for the 1994 High School Transcript Study were drawn from the pool of NAEP field supervisors. To avoid confusion, we refer to the data collection personnel for the HSTS simply as field workers. The field workers were trained in the data collection procedures for HSTS in December 1993. This training was conducted by the HSTS Curriculum Specialist/Coding Supervisor and took place over one full day. The training consisted of three sessions. The purpose of the first session was to establish the background knowledge needed to help field workers to make informed decisions when collecting information in the schools, and to explain why attention to detail and accuracy would be crucial in ensuring the quality of HSTS data. The second training session was held to familiarize field workers with the HSTS materials and forms and with the variety of materials they could expect to find in the schools. The third session provided an opportunity for field workers to work with sample catalogs and transcripts, and to fill out practice forms, as they would do using the actual materials for the HSTS. Exhibit 4-1 is a copy of the training agenda for the 1994 HSTS.

The first training session consisted of a presentation describing the purposes of the HSTS, the procedures Westat uses in handling and processing HSTS data, and the best sources of data to obtain from schools to provide Westat with the needed data.

During the second session, field supervisors were shown examples of various types of high school records and materials, including school- and district-level catalogs, course lists, and transcripts. The information on each of these materials was cross-referenced to the data needed for the HSTS at the school and student levels. Transparencies of screen prints of the transcript data entry and course coding systems were shown to demonstrate how the information from the specific materials would be used.

The third training session consisted of sets of exercises to complete to provide the field workers with hands-on experience in examining school materials and filling out the forms they would use. The practice materials consisted of copies of actual catalogs, course lists, and transcripts obtained in the 1990 HSTS (with all identifying information deleted).

Exhibit 4-1. Training agenda for the 1994 HSTS

1994 MAIN NAEP ASSESSMENT SUPERVISORS TRAINING SESSION

December 7 - 11, 1993

DAY 1 (Tuesday)

9:00 - 9:30	Introduction and Project Overview
9:30 - 10:00	Most Recent Data Releases (Press Conferences, Reports)
10:00 - 10:30	The 1994 Program - History of Contacts, Role of the Assessment Supervisor
10:30 - 10:45	Break
10:45 - 12:00	Supervisor's Assignment of Schools, Materials and Supplies
12:00 - 1:00	Lunch
1:00 - 3:00	Student Sample Selection and Preparation of the Administration Schedule
3:00 - 5:00	Field Managers Review Sampling with New Supervisor

DAY 2 (Wednesday)

9:00 - 12:00	Assessment Questionnaires Teacher Questionnaires IEP/LEP Student Questionnaires School Characteristics and Policy Questionnaires
12:00 - 1:00	Lunch
1:00 - 3:30	Presentation of Exercise Administrator Training Program to New Supervisors
3:30 - 3:45	Break
3:45 - 5:00	Classroom Management (Video)

Exhibit 4-1. Training agenda for the 1994 HSTS (continued)

1994 MAIN NAEP ASSESSMENT SUPERVISORS TRAINING SESSION

December 7 - 11, 1993

DAY 3 (Thursday)

9:00 - 10:00	Preparing for the Assessment Session
10:00 - 12:00	Conducting Assessment Sessions
12:00 - 1:00	Lunch
1:00 - 2:00	Concluding Sessions and Filling Out the Administration Schedule
2:00 - 3:00	Packing and Shipping
3:00 - 3:15	Break
3:15 - 5:00	Field Managers Review with New Supervisors

DAY 4 (Friday)

9:00 - 12:00	Transcript Study
12:00 - 1:00	Lunch
1:00 - 3:00	Transcript Study (continued)
3:00 - 5:00	Distribute Materials

DAY 5 (Saturday)

9:00 - 12:00	Field Managers meet with Supervisors to discuss administrative procedures, reporting, travel guidelines; and Scheduling Supervisors meeting with Assessment Supervisors to discuss schools and schedule
12:00 - 1:00	Lunch
1:00 - 3:00	Individual Study and Review

The first set of exercises was completed by the group as a whole, using transparencies of the materials and an overhead projector. The second set was completed in small groups, and the third set was completed individually and collected for review by supervisory staff. Errors or misconceptions were corrected and discussed with the field workers prior to their leaving the training session. Sample catalogs included a course list, extracts from a large catalog, and a smaller catalog. The sample materials were selected to give field workers a sense of the variety of materials they might expect to find in schools with respect to the amount of information available, the physical layout of the materials, and the ease or difficulty of accessing the information in the materials. Transcripts were examined in this exercise to show a number of ways that special education, for example, might be indicated, as well as indicators for transfer courses, remedial courses, honors courses, off-campus location courses, or courses for students with limited English proficiency.

4.2 Contacts with States, Districts, and Schools

In September 1993, superintendents and principals were notified about the transcript study through the Summary of School Tasks which was included in a mailout. This summary included information on several aspects of the main NAEP study, as well as the notification of the transcript study. In December 1993, district superintendents of participating 12th-grade schools sampled for the main NAEP and selected for the HSTS were mailed additional information concerning the HSTS. Items in the package included the following:

- An informational letter to school superintendents from Steve Gorman of NCES (Exhibit 4-2);
- A list of schools in the district selected for the 1994 HSTS; and
- A summary of school transcript activities (Exhibit 4-3).

For contacts with school-level personnel, field workers were provided with the following materials:

- An informational letter to principals from Steve Gorman of NCES (Exhibit 4-4);
- An informational letter to principals from Nancy Caldwell of NAEP/Westat (Exhibit 4-5); and
- A summary of school transcript activities (Exhibit 4-3).

Exhibit 4-2. Superintendent's letter from Steve Gorman

U S DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

NATIONAL CENTER FOR EDUCATION STATISTICS

January 1994

Dear Superintendent:

As described in previous mailings to your district, the 1994 High School Transcript Study is being conducted in conjunction with the 1994 National Assessment of Educational Progress (NAEP). The purpose of this study is to supply data to educational researchers and policy analysts on course-taking patterns and to examine the relationship of these patterns to achievement in secondary schools. The NAEP school sample is being used both because it is a nationally representative sample and in order that NAEP data and transcript data can be linked for schools participating in both. The participation of all selected schools (regardless of whether they are participating in NAEP) is needed to make the results of the transcript study comprehensive, accurate, and timely.

A list of the NAEP schools in your district selected for this study is enclosed. Detailed information on transcript activities in the school accompanies this letter. **No student time is involved.** Students' names and other individually identifying information will be removed from copies of the transcripts before they leave the school, and schools will be reimbursed at the standard rate for supplying transcripts.

Initial activities will be conducted at the same time NAEP supervisors are in the schools selecting the NAEP sample. In the fall of 1994, supervisors will return to the school to collect the requested transcripts.

The granting of Education Department authority for collection of the transcript data has been made pursuant to the provisions of the Family Education Rights and Privacy Act (FERPA) (20 U.S.C. 1232g), as implemented by 34 CFR 99.31(a)(3)(ii) and 99.35. These laws and regulations permit an educational agency to disclose records to authorized representatives of the Secretary of Education without the prior consent of the survey participants in connection with the audit and evaluation of Federal and State supported education programs. The privacy of the information schools are asked to supply to the NAEP contractors will be protected as required by FERPA and will be further protected by the removal of names and other identifying information. A copy of the relevant section of FERPA regulations is reproduced on the reverse side of this page.

I would appreciate your cooperation in this important component of the 1994 NAEP. If you have any questions about the study or its procedures, I may be contacted at the Department of Education or you may contact Nancy Caldwell of Westat, Inc., at (800) 283-6237.

Sincerely,

Steve Gorman

Steve Gorman
Project Officer

WASHINGTON, D C 20208—

Exhibit 4-3. Summary of school transcript activities

1994 HIGH SCHOOL TRANSCRIPT STUDY

SUMMARY OF SCHOOL ACTIVITIES

This sheet summarizes the High School Transcript Study activities that will be undertaken in 1994. Hopefully, it will provide answers to some of the questions you may have. NAEP Supervisors will provide you with a more detailed description of these tasks during telephone and in-person visits to the school.

KEY ASPECTS OF THE HIGH SCHOOL TRANSCRIPT STUDY

- NO STUDENT TIME IS INVOLVED. NAEP staff will work with your school and do as much of the work as possible to minimize the burden.
- Students' names and other individually identifying information will be removed from copies of the transcripts before they leave the school.
- Your school will be reimbursed at your usual rate for providing transcripts.

ACTIVITIES INVOLVING SCHOOLS

Phase 1: January - April 1994

1. The 1994 High School Transcript Study sample will be identified by the NAEP Supervisor.
2. Course lists or catalogs will also be requested. Course catalogs will be requested for the following years: 1993-94, 1992-93, 1991-92 and 1990-1991.
3. A sample of three transcripts will be requested. One should include regular courses, one special education course, and one honors course.
4. The NAEP Supervisor will need to review transcripts and course catalogs before leaving your school so that questions about either may be clarified.

Phase 2: Fall 1994

1. In the Fall of 1994, NAEP staff will return to your school to collect the requested transcripts of students who graduated.

Exhibit 4-4. Informational letter to principals from Steve Gorman

SAMPLE

January 1994

Dear Principal:

In conjunction with the 1994 National Assessment of Educational Progress (NAEP), the National Center for Education Statistics , U.S. Department of Education has authorized Westat, Inc., the NAEP contractor, to obtain student transcript data from a national sample of secondary schools sampled for the 1994 NAEP. The purpose of the 1994 High School Transcript Study is to supply data to educational researchers and policy analysts on course-taking patterns and the relationship of these patterns to student achievement in secondary schools across the nation.

Your school has been selected to participate in this important study and an informational letter has been sent to your District Superintendent. Your school's participation is needed to make the results of this study comprehensive, accurate, and timely. No student time is involved and schools will be reimbursed at the standard rate for supplying transcripts. Detailed information on the transcript activities and the timeframe for data collection accompanies this letter.

The granting of Education Department authority for collection of the transcript data has been made pursuant to the provisions of the Family Education Rights and Privacy Act (FERPA) (20 U.S.C. 1232g), as implemented by 34 CRF 99.31(a)(3)(ii) and 99.35. These laws and regulations permit an educational agency to disclose records to authorized representatives of the Secretary of Education without the prior consent of the survey participants in connection with the audit and evaluation of Federal and State supported education programs. The privacy of the information you are asked to supply to the NAEP contractors will be protected as required by FERPA, and will be further protected by the removal of names and other identifying information. A copy of the relevant section of FERPA regulations is reproduced on the reverse side of this page.

I would appreciate your cooperation in this most important component of the 1994 NAEP. If you have any questions about the study or its procedures, I may be contacted at the Department of Education or you may contact Nancy Caldwell of Westat, Inc., at (800)283-6237.

Sincerely,

Steve Gorman
Project Officer

Exhibit 4-5. Informational letter to principals from Nancy Caldwell

THE NATION'S REPORT CARD

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

1650 RESEARCH BOULEVARD • ROCKVILLE, MARYLAND 20850
TELEPHONE 1-800-283-6237 • FAX 301-294-2038

January 1994

Dear Principal:

Thank you for your participation in the 1994 National Assessment of Educational Progress. As indicated in the letter from Steve Gorman of the National Center for Education Statistics and as described in previous informational mailings regarding the 1994 national assessment, the U.S. Department of Education has authorized the National Assessment of Educational Progress (NAEP) to collect high school transcript data.

The purpose of this study is to obtain current information on course-taking patterns of high school students and to correlate this information with achievement data from the 1994 NAEP. To be nationally representative, the 1994 High School Transcript Study will include a sample of secondary schools selected for the 1994 National Assessment of Educational Progress. This is an important study and each participating school will make a valuable contribution to its success.

Detailed information on transcript activities in the school accompanies this letter. The activities for Phase 1 will be conducted at the same time that NAEP supervisors are in your school selecting the NAEP sample. Phase 2 of the study will occur in the fall of 1994 when the NAEP supervisor will return to your school to collect the requested transcripts. No student time is involved and schools will be reimbursed at the standard rate for supplying transcripts.

NAEP has been authorized to collect information on sampled students from their academic records pursuant to the provisions of the Family Education Rights and Privacy Act (FERPA). All students' names and other individually identifying information will be removed from the collected data before it is sent to our offices. All information obtained through this study will be kept confidential and will only be used for statistical reporting purposes.

Should you have any questions, please contact either me or Sandra Rieder at Westat (800) 283-6237.

Sincerely,

Nancy W. Caldwell

Nancy W. Caldwell
NAEP Project Director

Field workers provided these materials to the school principals and school coordinators during their initial visit to schools to conduct sampling. They discussed the HSTS with the school coordinator prior to the sampling visit when they called to confirm the sampling date.

4.3 Obtaining Course Catalogs, Sample Transcripts, and Other School-Level Information

Field workers requested sample materials for the HSTS when they first went to a school, and collected these materials when they returned to the school for sampling. The sample materials included a list of courses (preferably a catalog) offered for each of four consecutive years, from 1990 through 1994; a completed School Information Form (SIF) as shown in Appendix C; and three transcripts of students who graduated in 1993, representing a regular student, one with honors courses, and one with special education courses. Since these materials were unique to each school, receiving them before the collection of the actual transcripts enabled us to examine them and call a field worker or the school with any questions we had during the school year (i.e., before school personnel left for the summer). The catalogs and transcripts collected were also examined by the field worker who filled out a Course Catalog Checklist (Exhibit 4-6) and a Transcript Format Checklist (Exhibit 4-7) for each item collected and sent to Westat.

4.3.1 Catalogs

Our prior experience in coding course catalogs for previous HSTS studies led us to identify the following levels of priority for the type of catalog to request:

(1) a school-level catalog providing course titles and descriptions;

(2) a district-level catalog, if it indicated which courses were offered at the HSTS participating school;

(3) a course list by department that included general descriptions of course offerings by department;

Exhibit 4-6. Course catalog checklist

NAEP School ID: _____

Supervisor: _____

Course Catalog Checklist

Record each catalog title and check off **all** items which are identified in the course description materials you have collected.

	School Level Materials							
School Year	Catalog Title	Course Title	Course Number	Course Credits	Course Description	Course Level[1]	Special Codes[2]	Special Programs[3]
1990-91								
1991-92								
1992-93								
1993-94								

	District Level Materials							
School Year	Catalog Title	Course Title	Course Number	Course Credits	Course Description	Course Level[1]	Special Codes[2]	Where Offered[4]
1990-91								
1991-92								
1992-93								
1993-94								

1 - Identified as Regular, Honors, AP, Remedial, Special Education, ESL?
2 - Does the catalog describe what codes mean?
3 - Are Special Programs (Sp.Ed, IB, Vocational, etc.) included in this catalog?
4 - Does the district catalog identify courses offered at the sampled HSTS school?

Exhibit 4-7. Transcript format checklist

NAEP School ID _____

Supervisor _____

Transcript Format Checklist

Marked	Not Marked	Not on Transcript	
			1. Student's birthdate
			2. Student's race/ethnicity
			3. Student's gender
			4. Student's IEP/LEP status
			5. Student's graduation date
			6. Years attending this school
			7. Type of diploma awarded
			8. When a course was taken (year and semester)
			9. For a single course: a. course name b. number of credits awarded c. length of course (one year, semester, or other) d. grade received e. level of course (honors, remedial, SpEd, regular) f. transfer credit from another high school g. taught in another language (or ESL course) h. vocational courses i. location, if not taught at this school site
			10. Total number of credits received
			11. "Weighting" of course credits/grades (for honors or remedial levels)
			12. Are abbreviations or codes used on the transcripts? If so, indicate on the back of this form what they are and what they mean for those that are not obvious.

(4) a school-level course list without descriptions;

(5) a district-level catalog without any indication of which courses were offered in specific schools.

Field workers filled out a Course Catalog Checklist for the catalogs they obtained. This checklist served two purposes. First, it guided field workers in obtaining materials with the maximum amount of information possible that would be useful in the HSTS. Second, the checklist provided Westat staff with a quick way to review catalogs, so that they could request additional information if needed. Catalogs (or whatever material was available) were forwarded to Westat.

4.3.2 Sample Transcripts

Since transcript format varies greatly among school districts throughout the country, it was sometimes difficult to find where on a transcript the needed information was located. This, of course, presented an obstacle to uniform treatment of information on transcripts. Another difficulty was encountered in determining the meaning of "coded" information found on some transcripts, particularly codes indicating the level of courses -- that is, whether a course was honors or remedial level, or whether it was a special education course or part of another special program.

To solve this problem, we obtained sample transcripts of previous graduates, marked up to indicate where on the transcript the needed information was to be found, and how information regarding course level was coded. We requested three sample transcripts from each school: one containing honors level courses, one containing special education courses, and one "generic" transcript. Attached to each marked-up transcript was a Transcript Format Checklist, indicating the information to be marked, and whether or not that piece of information was included on the school's transcripts.

4.3.3 School Information Form

The School Information Form was forwarded to Westat along with the other preliminary materials as described above. The SIF (see Appendix C) was completed by either the field worker or a school staff member, or sometimes by both. The name and position of the school's HSTS coordinator

who helped fill out the SIF appeared on the first page. The completed SIF contained information about the school in general; about sources of information within the school, if needed to complete HSTS data collection; about the course description materials; about graduation requirements and grading practices at the school; and about the format of the school's transcripts. The field workers were instructed to fill out the SIF completely, or to indicate clearly on the SIF where the requested information could be found in the other materials provided by the school.

4.3.4 School Characteristics and Policies Questionnaire

The School Characteristics and Policies Questionnaire (SCPQ, Appendix B) is an 84-item questionnaire that collected information about school, teacher, and home factors that might relate to student achievement. It was completed by a school official (usually the principal) as part of NAEP for the NAEP participating schools. It was completed during the transcript data collection period for the remaining schools.

4.4 Identifying the Sample Students and Obtaining Transcripts

The HSTS used the NAEP sample for selecting schools and students. For schools that participated in NAEP, the student sample was recorded on the NAEP Administration Schedules. For schools that did not participate in NAEP, the field worker drew a sample of students at the school. Our procedures for identifying students in schools with NAEP materials and in schools without NAEP materials are described in detail in separate sections below.

4.4.1 Schools with NAEP Materials

Schools that participated in NAEP identified students participating in the HSTS at the same time that the NAEP sample was identified. For all HSTS participants, a brightly colored Disclosure Notice (Exhibit 4-8) was placed in the student's cumulative record folder where it would be highly visible and thus make it easier to identify and collect needed transcripts after students had graduated.

Exhibit 4-8. Disclosure notice

DISCLOSURE NOTICE

1994 HIGH SCHOOL TRANSCRIPT STUDY

Date: Spring Quarter 1994
Fall Quarter 1994

A copy of this student's transcript has been provided to WESTAT, Inc., agent for the U.S. Department of Education, National Center for Education Statistics (NCES). The granting of Education Department authority for collection of the transcript data has been made pursuant to the provisions of the Family Education Rights and Privacy Act (FERPA) (20 U.S.C. 1232g), as implemented by 34 CFR 99.31(a)(3)(ii) and 99.35. This disclosure statement fulfills the requirements of provision 34 CFR 99.32 of FERPA.

The High School Transcript Study (HSTS), sponsored by NCES, is being conducted to collect information on current course offerings and course taking in the nation's secondary schools. This student has been selected to participate in HSTS, and data from these records will be combined with others into statistical summaries and tables. No individually identifiable information will be released in any form.

For each NAEP school, the HSTS field worker was given a Transcript Request Form, Version 1 (TRF, Exhibit 4-9). The TRF was preprinted with information collected during the NAEP administration – specifically, each selected 12th-grader's NAEP ID Number, birth month, birth year, gender, and race. It also contained flags representing IEP, LEP, or Chapter 1 status. The field worker filled in the student name of each assessed, absent, or excluded student listed on the NAEP Administration Schedules.

The field worker obtained the student's exit status from the school staff and entered it in the Exit Status column using one of the following codes to describe each student's outcome at the school during this school year:

1. Standard diploma
2. Honors diploma
3. Diploma with special education adjustments
4. Certificate of attendance
5. Still enrolled in this school
6. Dropped out
7. Other (such as transferred, GED, unknown)

The following procedures for completing the Transcript Request Form were provided by the field worker.

1. Enter your name at the "Supervisor" line in the top box of the TRF.
2. Verify that the school has all of the pages of the Administration Schedules, comparing the school copies to your own. Students names should be legible on the complete, school copy.
3. Eliminate any non-twelfth graders by lining through their names. (A single line through the name will be sufficient.)
4. Begin with the NAEP ID of the first student on the Administration Schedule. Find the corresponding NAEP ID on the Transcript Report Form. (These are printed in ID order.)

1994 HIGH SCHOOL TRANSCRIPT STUDY
TRANSCRIPT REQUEST FORM FOR SCHOOLS COOPERATING IN NAEP AND IN TRANSCRIPT

SCHOOL ID:
SCHOOL NAME:
SUPERVISOR: _____

0 TRANSCRIPTS REQUESTED: _____
0 TRANSCRIPTS RECEIVED: _____

STUDENT NAME					COMPLETE IF MISSING						
FIRST	MI	LAST	NAEP ID	EXIT STATUS	SEX	BIRTHDATE	RACE/ETH	IEP	LEP	CHPI	TRANSCRIPT RECEIVED
			043-0815145		M	08/75	WHITE				
			044-0805401		F	06/76	WHITE				
			044-0816699		M	12/75	WHITE				
			045-0816210		F	07/76	WHITE				
			046-0807876		M	10/76	WHITE				
			047-0814930		F	06/76	WHITE				
			048-0812000		M	01/76	WHITE				
			048-0812166		M	09/76	BLACK				
			048-0812392		F	04/75	WHITE				
			049-0842882		F	03/75	WHITE				
			049-0842904		F	07/76	WHITE				
			049-0842912		M	12/74	WHITE				
			049-0842939		M	01/76	WHITE				
			049-0842955		M	04/75	HISP				
			049-0842963		F	04/76	WHITE				
			049-0842971		F	07/76	WHITE				
			049-0842980		F	01/76	WHITE				
			049-0843009		M	10/76	WHITE				
			049-0843013		M	10/76	WHITE				
			049-0843021		F	07/76	BLACK				

Exhibit 4-9. Transcript request form - Version 1

5. The birth date, sex, race/ethnicity, IEP/LEP and Chapter I status, should all be pre-printed on the TRF and should match the information recorded for that student on the Administration Schedule. If not, correct the information on the TRF, after you have verified that you have matched entries correctly.

6. Record the student's full name from the Administration Schedule on the line of the Transcript Request Form with the same NAEP ID. Make a small check on the Administration Schedule as you go to indicate you have completed the transcription for a given student (this should be the last use of the Administration Schedule). In some schools, it may be necessary to record some form of school ID (e.g., Social Security Number) in addition to or in lieu of the student's name for the school to access their files. Make sure you're aware of this before you start completing the TRFs.

7. Continue this process for all twelfth-grade students on the Administration Schedules with one exception: any students who have been crossed-off as "withdrawn" should be skipped in the process.

8. When you have gone through all of the Administration Schedules in this fashion, you should have a name entry corresponding to each NAEP ID pre-printed on the TRF.

9. The "exit status" for each student may be coded at this time if it is available. Alternatively, this information may be recorded when the transcripts are received. Confer with your School Coordinator to determine the best way to get this information; it may not be on the transcript or it may be coded information.

10. Record the number of transcripts requested in the box at the top of the first page of the TRF. Record the number received at the time you obtain the transcripts. For each transcript received, place a checkmark in the "Transcript Received" column. Be sure to complete a "Documentation of Missing Transcripts" form (Exhibit 4-10) if you cannot obtain a transcript.

Once the field worker filled in the names of the students, most schools were generally able to obtain a data file and copy the transcripts. In other schools, the transcripts were pulled from their folders and photocopied at the school.

Once the request was filled, the field worker reviewed the transcripts to ensure that she received a transcript for each 12th-grade student who was selected for the NAEP assessment, whether or not that student had graduated. The field worker then checked each transcript for eligibility, understandability (e.g., are all the codes on it defined on the transcript or explained in the SIF?), and

Exhibit 4-10. Documentation of missing transcripts

DOCUMENTATION OF MISSING TRANSCRIPTS

School Name: _____ Date: _____

School ID #: _____

Supervisor: _____

Number of Transcripts Requested: _____

Number of Transcripts Received: _____

Reason(s) School Gave for Missing Transcripts:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

completeness and labeled each transcript with preprinted labels containing the School ID and the NAEP ID for each student. The field worker completed a "Documentation of Missing Transcripts" (Exhibit 4-10) form to explain the reasons the school gave for any missing transcripts.

After the field worker collected and recorded all the information required on the sampled students and reviewed the transcripts for completeness and accuracy, he or she prepared the transcripts for transmittal to Westat. This procedure involved "masking" all personally identifiable information where it appeared on each transcript, using a broad felt tip marker or correction tape to line through or cover all identifiers.

ersonal identifiers were also removed from the Transcript Request Forms. Before returning the TRFs to Westat, the field worker cut off the portion that included the students' names, in order to comply with our confidentiality provisions. The portion with the names was left in the school's NAEP folder.

Schools were reimbursed for the transcripts at their standard rates. The field worker then completed a Shipping Transmittal Form (Exhibit 4-11) and returned it with the TRF, the transcripts, the Documentation of Missing Transcripts, and the SIF to Westat.

4.4.2 Schools without NAEP Materials

In schools that did not participate in NAEP, the field worker first selected a sample of students. She then requested transcripts for those students and followed the procedures described in the previous section for reviewing and shipping transcripts. She also completed the School Information Form, requested that the SCPQ be completed, and collected course catalogs for the past four academic school years (1990-91, 91-92, 92-93, and 93-94). The information included in the catalogs was documented by completing the Course Catalog Checklist. At this point, the procedure was different. Rather than obtaining and annotating three example transcripts, as was done at the time of the NAEP visit to the school, the field worker used the Transcript Format Checklist to annotate the first transcript she collected.

Exhibit 4-11. Shipping transmittal form

908842

1994 HSTS - SHIPPING TRANSMITTAL FORM

(INSTRUCTIONS: Fill out for each school and shipment)

School ID #: _____ School Name: _____

Supervisor: _____ School Shipment #: 1 2

Date Shipped: _____ Source of Sample: ☐ NAEP List
☐ New Sample

1. TRANSCRIPTS:

1) Total Number Requested _____
2) Number in This Shipment _____
3) Number Unavailable _____
4) To be Sent _____

2. IEP/LEP STUDENT QUESTIONNAIRES:

1) Total Number Requested _____
2) Number in This Shipment _____
3) Number Unavailable _____
4) To be Sent _____

IF SCHOOL DID NOT PARTICIPATE IN NAEP, COMPLETE THE FOLLOWING TWO QUESTIONS.

3. COURSE CATALOG: (check one)

☐ In This Shipment
☐ To be Shipped
☐ Unavailable

4. SCHOOL INFORMATION FORM: (check one)

☐ In This Shipment
☐ To be Shipped

In the schools that participated in HSTS but not in NAEP, the process of generating a sample of students began when the school produced a listing of all students who graduated from the 12th grade during the spring or summer of 1994. This list was requested during the preliminary call placed to the school when it was determined that the school would participate in HSTS. The following information was collected for each student selected for participation in HSTS:

- Exit status,
- Sex,
- Birthdate (month/year),
- Race/ethnicity,
- If Individualized Educational Program (IEP),
- If Limited English Proficiency (LEP),
- If receiving Chapter I services.

These data were collected either with the list of 1994 graduates or after sampling, depending on which procedure was easier for the school.

Selecting the Sample

As already noted in Section 3.2, there were two basic sampling rules for the 1994 HSTS. These rules applied to all schools that required a new sample of students.

1. If there were 60 or fewer graduates listed, all were included in the sample.
2. If there were more than 60 graduates listed, a sample of 50 students was drawn using a systematic sample.

Because the students in the HSTS schools did not have NAEP identification numbers, a set of IDs was preassigned for up to 60 students in each school. The Transcript Request Form--Version -- 2 (Exhibit 4-12) was preprinted with these IDs and had space for filling in each student's name and basic demographic characteristics.

1994 HIGH SCHOOL TRANSCRIPT STUDY
TRANSCRIPT REQUEST FORM -SCHOOLS COOPERATING TRANSCRIPT AND REFUSING NAEP

SCHOOL ID:
SCHOOL NAME:
SUPERVISOR: _____

TRANSCRIPTS REQUESTED: _____
TRANSCRIPTS RECEIVED: _____

STUDENT NAME			TRANSCRIPT STUDY ID	COMPLETE IF MISSING:							TRANSCRIPT RECEIVED
FIRST	MI	LAST		EXIT STAT	SEX	BIRTHDATE	RACE/ETH	IEP	LEP	CHP1	
			990-0000101								
			990-0000102								
			990-0000103								
			990-0000104								
			990-0000105								
			990-0000106								
			990-0000107								
			990-0000108								
			990-0000109								
			990-0000110								
			990-0000111								
			990-0000112								
			990-0000113								
			990-0000114								
			990-0000115								
			990-0000116								
			990-0000117								
			990-0000118								
			990-0000119								
			990-0000120								

Exhibit 4-12. Transcript request form - Version 2

The field worker, with the assistance of the school, completed the TRF and submitted it to the school staff. The transcripts were then received by the field worker, reviewed, and shipped to Westat in the same manner as transcripts from schools participating in NAEP.

4.5 IEP/LEP Questionnaire

One of the questionnaires obtained in the HSTS was the IEP/LEP Questionnaire. This was completed for students for whom the school had developed an Individualized Educational Program (IEP) and for students with Limited English Proficiency (LEP). We asked the schools to have the person most knowledgeable about a student complete the IEP/LEP questionnaire. In large schools, this person was typically a counselor, a special education teacher, or a teacher of English as a Second Language. In smaller schools, this person was typically a classroom teacher.

The questionnaire was completed according to the program in which the student was enrolled. Question 1 ("Why is this student classified IEP/LEP?") and Part A (questions 2 through 4) of the questionnaire were completed for both groups of students (i.e., those classified as disabled and for those classified as having limited English proficiency. Part B of the questionnaire (questions 5 through 14) was completed only for students with an IEP (i.e., students with disabilities). Part C (questions 15 through 26) was completed only for students with limited English proficiency. A copy of the questionnaire is included as Appendix D.

For schools participating in the 1994 NAEP, the IEP/LEP questionnaires were collected as part of the NAEP procedures. In schools with newly sampled students, the school identified the IEP/LEP students in the sample and filled out the questionnaire for each student.

Identical IEP/LEP questionnaires were used for NAEP and HSTS. The IEP/LEP forms collected during NAEP were scanned by National Computer Systems (NCS) and the file provided to Educational Testing Service (ETS). ETS provided Westat with data for all 12th-grade students (N=2,472) for whom the IEP/LEP questionnaires had been completed during NAEP. Another 69 IEP/LEP questionnaires were collected during the HSTS and scanned by NCS using the same procedures as were used for the NAEP IEP/LEP questionnaires. NCS forwarded this data file directly to Westat. Of these questionnaires, only the ones with corresponding records in the Student File were selected for the final IEP/LEP file. A total of 2,541 students are represented in the final IEP/LEP file.

4.6 Sending Data to Westat

As with NAEP, safeguards were built into the procedures for the transcript study to ensure that applicable privacy requirements were met. These safeguards included the removal of all personal identifiers from the transcripts provided by the schools. When the transcripts left the school, students could be identified only by ID numbers. In schools where the NAEP information was available, the ID number was the same as the student's NAEP booklet number. In schools where a sample of students was drawn, new IDs were generated.

After transcripts were collected and all information on sampled students recorded, field workers prepared the transcripts for transmittal to Westat. They first compared the data on the transcripts to the TRF to verify that they had obtained and correctly labeled the transcripts. At the same time, they noted on the TRF which transcripts were received and which were not. They then used scissors to cut off the left hand column of the TRF, which contained the names of the students. The list of names was destroyed and the remainder of the TRF was placed in the package to send to Westat.

The field workers masked all personally identifying information where it appeared on each transcript, using a broad felt tip marker to line through all identifiers. The types of personal identifiers and their location on the transcripts were different for each school and, sometimes, were different for the different categories of students within a single school. Field workers were careful to examine every transcript and line through the following information each time it appeared: student's name, parent's name, names of guardians or other relatives, addresses (including street, city, state, ZIP), and phone numbers.

A Shipping Transmittal Form accompanied all shipments to Westat and summarized the types and number of materials being sent. This form also gave information on whether the transcripts were from the NAEP list or a new sample and, if the school did not participate in NAEP, whether course catalogs and SIFs were included in the shipment.

4.7 Receipt and Review of Data from Data Collectors

When transcript study materials arrived at Westat, a receipt clerk carefully reviewed all items for accuracy and completeness. Transcripts were matched to the Transcript Request Form. Field workers were contacted immediately if further clarification was needed. Schools were reimbursed for the cost of producing the transcripts within 2 weeks of having their materials received at Westat.

An automated management system was developed and maintained at Westat. A disposition code structure was developed to indicate the status of each school's participation. As field workers reported the results of their contacts with district superintendents and individual schools, a receipt clerk keyed a disposition code for each school. Disposition reports were generated from the receipt system once a week so that home office staff could review the progress of securing cooperation from the sampled schools. Overall, the cooperation rate was 90 percent. Of the 379 schools sampled for the HSTS, 340 agreed to participate. Of the 340 schools, 283 also participated in NAEP, while 57 refused to participate in NAEP.

Once verified, information on the number of transcripts and course catalogs requested and received was entered in the receipt system by a data entry clerk. Weekly status reports were generated to monitor the progress of obtaining the transcripts. Transcripts and other school materials were maintained in individual school folders and stored until used by data preparation staff.

Catalogs, sample transcripts, and SIF's were reviewed at Westat to ensure their completeness. Phone calls were made to the field workers or to schools, as needed, to resolve any questions regarding the content or accuracy of the materials.

5. DATA PROCESSING PROCEDURES

When entering and cleaning the data for the 1994 NAEP High School Transcript Study, we performed the following steps:

- Establishing student ID control lists;
- Entering transcript data;
- Coding the catalogs;
- Matching transcript titles to catalog titles;
- Standardizing credits and grades;
- Quality control;
- Scanning and preparing the IEP/LEP questionnaires; and
- Scanning and preparing the School Characteristics and Policy questionnaires.

The first six steps are closely related and involve the entry and coding of the students' transcripts and the schools' catalogs, as well as matching the courses on the coded catalogs to the courses on the transcripts. The last two steps were actually performed in parallel with each other and the first six. They involve the data entry and formatting of data provided to us on optical scan forms by school personnel.

Each of the steps is described in detail in a separate section below.

5.1 Establishing Student ID Control Lists

Student ID control lists were developed from lists obtained from the NAEP administration records for schools that participated in NAEP. The control list for a school is the master list of IDs against which all other operations are checked. Only IDs matching those on the control lists are processed, as other IDs are either out of scope or miskeyings. In addition, each data processing step must account for all the IDs on the control list or for a well-defined subset of those IDs. Only NAEP students who were identified during the NAEP administration as 12th graders were retained on the control lists generated from NAEP.

Students identified as 10th or 11th graders, or those with an unknown grade, were removed from the ID control lists. For schools that did not participate in NAEP, or had lost the linkage between the student's names and their IDs (only three schools), control lists were compiled from completed transmittal request forms. A data file was created for each school listing the valid student IDs for that school.

5.2 Entering Transcript Data

Transcript data entry began in June 1994, as transcripts were received from the schools. Data entry personnel entered transcript data using a Computer Assisted Data Entry (CADE) system. The system displays labeled blank fields and the data entry clerk fills in the fields. It checks each entry to verify that it is within an allowed range and warns the clerk when a problem occurs. The coding supervisor conducted 2 days of training, consisting of instruction in the use of the CADE system for data entry and interpretation of the extensive variety of formats found in the transcripts.

Data entry clerks were instructed to use the Transcript Format Checklist (see Exhibit 4-7) as a source of information. The checklist included student's birthdate, race/ethnicity and gender, IEP/LEP status, graduation, type of diploma awarded, details about an individual course, total number of credits received and whether abbreviations or codes were used on the transcript.

We used actual transcripts illustrating different formats and different types of information as demonstration materials. Trainees also used these transcripts as practice exercises to gain familiarity and skill in using the CADE system.

In addition, two experienced HSTS data coders prepared a summary sheet for each school which directed the data entry clerk's attention to any special features or difficulties associated with a set of transcripts.

CADE System

The CADE system included three basic data entry screens. The first screen was used to enter student-level information (date of birth, date of graduation, type of diploma, attendance information, grade point average, and class rank). The second was used to enter data on honors and scores on standardized

tests. The third screen was used to enter course data from the transcripts, including course title, grades, credits received, year taken, and a number of "flags" indicating whether a course was a transfer course, an off-campus course, an honors course, a remedial course, or an ESL course (or taught in a foreign language). The data for all the students in one school were collected in a set of three database files, one file corresponding to each of the three screens.

Data Entry Procedures

Transcript data entry clerks using CADE (referred to as CADErs) selected a school and began entering each eligible transcript (transcripts for students who did not graduate or who were deemed out of scope were not entered) for that school, with each CADEr working on one school at a time. They entered data exactly as it appeared on the transcript, except that they were instructed to use abbreviations as indicated in Exhibit 5-1 and to change all Roman numerals to Arabic numerals. We instructed all CADErs to direct any questions or problems to the curriculum specialist or to one of the experienced data coders. When all transcripts for a school were completed, the status of the school file changed from "incomplete" to "ready for verification."

Exhibit 5-1. Abbreviations for data entry

Advanced.....Adv	Honors.....Hon
Advanced Placement.....AP	Industrial Arts.....IA
American.....Amer	Intermediate.....Intermed
Beginning.....Beg	International Baccalaureate.....IB
Biology.....Bio	Introduction.....Intro
College Prep(aratory).....CP	Mathematics.....Math
Cooperative.....Coop	Physical Education.....PE
Education.....Ed	Science.....Sci
English.....Engl	Special Education.....SpEd
General.....Gen	Trigonometry.....Trig
Government.....Govt	United States.....US
History.....Hist	Vocational.....Voc

Verification of Data

All transcript data was 100 percent verified by a CADEr other than the one who initially entered the data. The verification portion of the CADE system is essentially a "re-do and match" process where data are re-entered (blind to the first entry), and the computer stops when a non-match between the original data and the current data is encountered. Verifiers can then either accept the original entry or override it with the verified entry.

All fields were rekeyed except the course name field, test name field, and honors name field. These three fields were displayed and reviewed by verifiers but were not key verified. As the three "name" fields were not used for any automated analyses and required the greatest number of key strokes to enter, it was felt that the most cost effective use of resources was to perform a visual verification rather than a rekeying. In addition, allowing the verifier to see the name of the course, test, or honors being entered greatly simplified the task of ensuring that the verifier entered data in the same sequence as the original keyer.

5.3 Coding the Catalogs

Catalog coding was performed by a staff of trained coders, all of whom had prior experience teaching. Two of the HSTS coders, who had served in this role in the 1990 HSTS, acted as task leaders on the 1994 HSTS.

Training of HSTS catalog coders took place over a 4-day period, where coders were trained in the catalog coding task and in the use of the computer system which they used to perform the coding process. The curriculum specialist conducted the training, using sample materials from the 1990 HSTS.

5.3.1 Course Title Entry

Titles of courses offered at each school included in the HSTS were entered from a catalog of course offerings provided by the school.[11] For the 22 schools that provided no listing of their courses, a

[11] In some cases, this was a district-level catalog. See Chapter 4 for a discussion of catalog types.

course list was created for the school, based on all the course titles appearing on students' transcripts (excluding courses that were transferred from other schools).

A curriculum specialist examined all catalog listings, regardless of how the catalog was created. Every attempt was made to eliminate duplicates and to ensure that course titles included appropriate annotations for grade ("English 10"), level ("Biology, AP"), or special programs ("Automechanics Coop Ed"). Errors were corrected by data entry personnel and the corrected list was again reviewed by the curriculum specialist.

Two variables indicating the source of information for a given school's catalog are provided with the School File. One variable indicates whether or not the course list that we used was derived from transcripts. The other indicates the type of catalog which the school provided (none, district catalog, school-specific course list, or school-specific catalog). For ease of use, these variables also appear in the Course Offerings File.[12]

5.3.1.1 School-level Catalogs or Course Lists

If a school provided a catalog of course offerings (as requested), data entry personnel entered a list of all course titles appearing in the catalog.[13] We made a concerted effort to standardize the format of titles. We converted all Roman numerals to Arabic numerals. We also standardized abbreviations of frequently appearing courses (or words in courses) such as "ADV" for "advanced," or "BEG" for "beginning," or "INTRO" for "introduction." These abbreviations are the same as those used by the transcript data entry clerks (see Exhibit 5-1).

About half of the schools that provided course catalogs provided one catalog representing the 1993-94 school year. Usually the School Information Form (see Section 4.3.3 and Appendix C) indicated that there had been no significant changes in course offerings over the 4 years in which graduating students attended the school. If a school provided more than 1 year's catalog, we evaluated them all to determine whether there were significant changes over the years provided. If we looked at a large number of courses

[12] A short description of each public use file created by the project is provided in Chapter 7.

[13] School-level course catalogs were provided by 196 schools. Another 78 schools provide school-specific course lists.

and determined that there were few differences across the catalogs, we entered the one for the 1993-1994 school year.

If more than one catalog was provided and they differed significantly among the years they covered, titles from more than one catalog were used. A curriculum specialist selected the portions of each catalog to be used so that they excluded sections on programs that students could take only by attending another school in the district, courses taken at night, and so on. The specialist included programs from previous years that were not listed in the current catalog but were offered during the period when students in the HSTS attended the school. These titles were then entered in the order of their appearance in the catalogs.

When we encountered a transcript course that was not a transferred course and did not appear in the 1993-1994 catalog, we examined previous catalogs to find a description of the course, if it was available, so that it could be appropriately coded. Whether or not such a course appeared in the catalogs, we added it to the Course Offerings File.

5.3.1.2 District-Level Catalogs

We found both school-level and district-level catalogs at the schools. Forty-four schools provided catalogs of courses offered by their entire school district, while the individual school's specific course offerings were a subset of those included in the district catalog. Often these district catalogs (which were quite large) included programs that we know are not offered at the school (such as an International Baccalaureate program, a vocational program, or a performing arts program). To create a listing of courses actually offered at such schools, we created a list in the same manner as for schools not providing any catalog (i.e., creating it from titles appearing on transcripts), but supplemented the resulting list with courses from the district catalog that were likely to be offered in the HSTS school (such as Advanced Placement English 12, Accounting, or Basic Biology) even if they did not appear on a transcript. Thus, the Course Offering File represents our best approximation to the complete list of courses offered by their schools to the 1994 graduates in our sample.

5.3.1.3 Schools without Catalogs

Approximately 6.5 percent of the schools (22 of 340 schools) did not provide any list of courses offered at the school. For these schools, which were most often very small, a course list was generated during the process of transcript data entry. When a course was entered that did not already appear on a course offering list, it was added to the list using a function key, which was programmed specifically for this purpose. The resulting list of courses taken by students at the school was then treated as the school's catalog.

There are significant limitations of creating catalogs for a school in this manner: (1) the list represents only courses taken by students in the sample, and may not include all courses actually offered at that school; (2) many courses are repeated, since the same course may have been entered into the transcript file in two different formats (e.g., "CONSTRUCTION 1" and "CONSTRUCTION TRADES 1, "or "GLBL STDY 9" and "GLOBAL STUDIES 9"), and (3) no course description is available to clarify the meaning of a title. These catalogs required considerable review and editing before course coding could proceed.

5.3.2 Course Coding

Course coding is the process of associating a course title with a classification code and setting a group of flags appropriately. The process involves selecting a course description from the classification system that most closely matches the course description in the course catalog.

5.3.2.1 Classification of Secondary School Courses

We used the *Classification of Secondary School Courses* (CSSC), including modifications we made during the 1987 and 1990 HSTS studies, as a standard for classifying and coding the courses offered by all the schools included in the HSTS and for classifying and coding all courses appearing on transcripts of students included in the HSTS. The CSSC is a six-digit, hierarchical numbering system for all regular and special education courses offered in American secondary schools. Each CSSC entry includes a six-digit code, a course title and alternate titles, and a course description. Westat updated the CSSC significantly in 1989 to reflect the changes we found in the breadth and types of courses taken by

students in the 1987 HSTS. We again supplemented the CSSC in 1992 by adding 14 new courses encountered during the 1990 HSTS. Appendix E lists the 12 courses that we added to the CSSC for the 1994 HSTS. No existing CSSC courses were deleted, nor were any existing codes changed.

5.3.2.2 Flags

We coded additional information for each course as a series of single-digit "flags." These flags were used to indicate special features of a course such as its relationship to other courses in a sequence of courses, the language of instruction for the course, the level of the course (honors, regular, or remedial), whether it was a combination course (a multi-subject course requiring multiple codes such as an art appreciation/music appreciation course), the location at which the course was taught, and any enrollment restrictions (regular or handicapped students). A full list of flags and their values is shown in Exhibit 5-2.

Exhibit 5-2. Values for flags

Sequence Flag:

- 0 Non sequential course (Default)
- 1 First course in sequence
- 2 Advanced course in sequence

Language Flag:

- 0 Taught in English (Default)
- 1 Taught in language other than English

Remedial/Honors Flag:

- 1 Honors course
- 2 Regular course (Default)
- 3 Remedial course

Off Campus Flag:

- 0 No (Default)
- 1 Yes, at area Vo-Tech
- 2 Yes, at Special Ed Center
- 3 Yes, other
- 4 Yes, at multiple locations

Combination Course Flag:*

- 1 Not a combination course (Default)
- 2 The course was assigned 2 CSSC codes
- 3 The course was assigned 3 CSSC codes
- 4 The course was assigned 4 CSSC codes

Transfer:

- 0 Not a transfer course (Default)
- 1 Transfer course

Handicapped:

- 0 Self contained special education
- 1 Non special education (Default)
- 2 Resource special education

* A combination flag was set when we needed to assign multiple CSSC codes to a course. When this happened, the course title was repeated, the course credits were divided evenly among each of the codes, and the combination course flag was set for each occurrence of the course title. A distinct CSSC code was then assigned to each occurrence.

5.3.2.3 Training Course Coders

Course coders who worked on this study had to meet a level of proficiency that would ensure accurate and high-quality results. We selected catalog coders who (1) had current or prior experience teaching in American schools and/or (2) had a college degree in education. An expert in special education was selected to code the special education courses for all schools. Two of the catalog coders had coded catalogs during the 1990 HSTS and were highly experienced. They assisted in part of the training and performed some specialized functions throughout the process of coding catalogs and entering transcript data.

Coder training was conducted over a 4-day period by the curriculum specialist, who was also the coding supervisor. Coders were trained both in the analytic aspects of selecting the best CSSC code for each course and in operating the CACE system. Training materials included practice exercises based on actual catalogs and transcripts from HSTS schools. The first day of training consisted of classroom-type presentation and a demonstration of the CACE system. The second day started with directed hands-on practice using CACE with training materials, and gradually moved toward more independent use of the system. On the third day, coders began working in pairs, using CACE to code their first actual catalog. Each coder's understanding of the coding task and CACE operation was evaluated each half-day on practice tests and exercises. The final day was devoted to the beginning of actual coding, but all work was carefully reviewed before it was considered complete.

All coders performed 90 percent or better on each evaluation before training progressed to the next stage. Additional training was conducted as needed when there were changes in the software or personnel. We also trained the catalog coders to use CACE to match transcript titles to course titles in catalogs.

5.3.2.4 CACE System for Catalog Coding

The Computer Assisted Coding and Editing (CACE) is a Paradox-based system that we designed specifically for coding high school catalogs. It consists of two major components: (1) a component for selecting and entering the most appropriate CSSC code and "flags" for each course in a catalog and (2) a component for matching each entry appearing on a transcript with an entry in the corresponding school's list of course offerings. In addition to providing for data selection and entry, CACE

maintains file consistency and produces output files suitable for further analysis and manipulation. CACE's user interface is designed to reduce the likelihood of user errors by encouraging selection from a list rather than key entry of necessary data items.

For the HSTS, the CACE system presented each title in a school's catalog to the catalog coder one at a time. The catalog coder then examined a "suggestion list" of potential codes for that course. The list consisted of CSSC codes that were assigned to similar titles in the 1990 HSTS. The list was synchronized with an on-line version of the CSSC (in another window on the same screen) so that the coder could compare the description for the course in the CSSC with the description in the school catalog. The coder selected the appropriate CSSC code either in the suggestion list or in the corresponding section of the CSSC. Alternatively, the coder could type the CSSC code directly into the appropriate data field on the screen.

The CACE system checked all entries against the master CSSC list before allowing the record to be stored in the database. If the items in the suggestion list were not good matches to the course description, the catalog coder could browse through the full on-line CSSC or look in the hard copy of the CSSC provided to each coder. If the coder could not determine an appropriate code for a course, he or she could select a special code from the suggestion list that marked the course for further consideration by the coding supervisor.

Codes for flags (described in Section 5.3 2.2) were automatically set to default values when a course was selected or entered and could then be changed to non-default values by the coder. The CACE system also included a "browse" screen where the catalog coder could rapidly review the work but could not edit it. This screen displayed the data using one line per course title, a format that particularly useful for locating uncoded entries and reviewing similar titles for consistency.

5.3.2.5 Catalog Coding Principles and Procedures

To assure consistency and quality, we based catalog coding decisions on a basic set of coding principles and procedures. First, the catalog coder reviewed a school catalog "holistically" to ascertain ways that course levels, special education, and other special programs were designated. He or she looked for sequences of courses, descriptions of programs, requirements, credits awarded, or other information provided, to obtain a general view of the curriculum. Then, using CACE, the coder looked at each course

title, found it in the catalog, and read whatever description was available. The coder then selected the best CSSC code for the course. Wherever possible, the database coder selected codes based on a course description rather than on title.

After selecting the CSSC code, the coder reviewed the flags for that course and edited them as needed. If the coder found courses in the CACE catalog listing that should not be there, they could be deleted. Similarly, if the coder found that a course was missing from the CACE listing of catalog titles, it was added to the list and coded. After the coder finished coding the regular education courses for a school, the special education expert coded all special education courses.

5.3.2.6 Coding Transfer Courses

An important variation on the course coding procedure was for transfer courses -- that is, those courses on a student's transcript that were taken when the student attended another school (but the credits for these courses were transferred to the HSTS school and accepted there). These courses were automatically added to the catalog list appearing in CACE with the "transfer flag" indicating their transfer status. In coding these transfer courses, the catalog coder could use only the course title to assign CSSC codes. No descriptive information was available unless the transfer course was taken in the same school district and we had a district catalog.

To address the issue of transfer courses, the CACE system built a list of transfer course titles and previously assigned CSSC codes and used these to assign CSSC codes automatically to transfer courses that matched items in the list. When a new transfer course was coded, it was added to the list. Since the number of transfer titles for a school could be quite large -- sometimes up to 80 percent of the titles for the entire school -- this automated procedure saved a great deal of time and ensured that identical titles always received identical codes.

5.3.2.7 Coding Special Education Courses

All special education courses were coded by a specialist holding a doctorate in special education. All special education coding was also reviewed by the curriculum specialist, who has extensive

expertise in special education. Special education courses were coded using the same procedures and CACE features as those used for other courses.

5.4 Matching Transcript Titles to Catalog Titles

Catalog coders completed a table that associated each course title appearing on a transcript with the title of a course in the school's catalog and its corresponding CSSC code and flags. The process was somewhat more difficult than might be expected because of the lack of uniformity in how courses were entered on transcripts, even within the same school. The task was also somewhat complex because flags as well as course titles must be matched, so that "Algebra 1" with an honors flag was appropriately matched with an honors level course in the catalog. For all schools, special education titles on transcripts were matched to appropriate catalog titles by the specialist in special education.

5.4.1 CACE System for Matching Titles

The CACE system includes a facility for matching titles of courses appearing on one or more transcripts in a school to a course appearing in the course catalog. When a catalog coder entered the title matching facility, the system divided the screen into two windows. The upper window contained a scrollable list of transcript courses in alphabetical order and their associated transfer flag, language flag, and remedial/honors flag. The lower window contained a scrollable list of course titles from the high school's catalog and their associated flags. The catalog coder selected a course title in the upper window and then scrolled through the list in the lower window to find the matching catalog title. The coder specified the matching catalog course by highlighting it and pressing the Enter key. The catalog title then appeared next to the corresponding transcript title in the upper window. This process continued until each transcript title was associated with a catalog title. To minimize the effort required for title matching, each transcript title was presented for matching only once. Thus, even though "English 9" appeared on all the transcripts from a school, the coder needed to match it only once.

Coders performed manual title matching only for non-transfer courses. Transfer titles were automatically matched by CACE since the catalog entries are copies of transcript titles. For transfer courses, a copy of the title of each transfer course was placed in the catalog course listing file so that it

could be coded with an appropriate CSSC code. Since these titles in the catalog are identical to those appearing in the transcript course list, they could be matched to one another automatically.

After all unique course titles on the transcripts were matched with catalog titles, and hence with their CSSC codes, a batch process used the matching information to automatically associate the appropriate CSSC codes with each transcript title.

5.4.2 Transcript-Catalog Association Principles and Procedures

We assigned a CSSC code to each course listed on a transcript by matching each unique course title on a transcript to a specific CSSC-coded course in the school's catalog. The CSSC code thereby was associated with the transcript title. The associations were based on a match of the title, level (i.e., average, honors, remedial), and flags (transfer, language of instruction, disability) for each transcript entry. The matching process also serves as an additional check on the accuracy of both transcript and catalog title data entry. For example, if an entry appears in the transcript but not in the catalog, the catalog coder reviews the transcript to determine whether the course should actually have been marked with the transfer flag. The coder reviews the catalog to determine whether the course was erroneously omitted from the list of catalog titles. Sometimes this process revealed entire programs that students took that were not described or even mentioned in the school catalog. This discrepancy may have occurred because the only catalog provided to us was out of date and different courses were offered in 1990-1994 than are represented in the older catalog.

One of the major difficulties we encountered in evaluating transcript course titles occurred when course titles were abbreviated. The original meaning of these abbreviations was difficult to determine. Some abbreviations could be deciphered by knowing the program offered at a school (e.g., "EFE" is "Economics and Free Enterprise"), but others remained indecipherable, despite all of our efforts (e.g., "ARCS"). Some titles could reasonably be assigned to a broad domain, if not a specific course. For example, "ABC Math" can be matched to the "Math-Other" course title and CSSC code. We matched an ambiguous title to an "other" course and code within a specific discipline whenever possible; otherwise the course was assigned a code of "600000," which means "uncodeable." This code was assigned to 706 of the over 1,000,000 courses entered. It represents less the 0.1 percent of the transcript entries.

5.5 Standardizing Credits and Grades

Since credit and grade information reported on transcripts varied considerably among schools, districts and states, it was necessary to standardize this information so that valid student-level and school-level comparisons can be made. We standardized credit information based on the Carnegie Unit, which we defined as the number of credits a student received for a course taken every day, one period per day, for a full school year. For each school, the catalog coder filled out a "Carnegie Unit Report" (as shown in Exhibit 5-3). The factor for converting credits reported on the transcript to the standard Carnegie Unit was verified by the curriculum specialist and then key entered for each school by data entry personnel.

Grade information on transcripts varied even more widely than credit information. Grades were reported as letters, numbers, or other symbols on a variety of scales. Coders provided standardized information for each school using the form shown in Exhibit 5-4 ("Standardization of Grades"), which were then key entered for each school by data entry personnel. Numeric grades were converted to standardized grades as shown in Table 5-1, unless the school documents specified other letter grade equivalents for numeric grades.

Table 5-1. Numeric grade conversion

Numeric grade	Standard grade
90-100	02 = A
80-89	05 = B
70-79	08 = C
60-69	11 = D
<60	13 = F

5.6 Quality Control

Each stage of the process described above included measures to assure both the quality and consistency of the data. Quality control (QC) procedures ranged from those for specific data items to those for a broad overview of the data. We describe these in more detail in the following sections.

Exhibit 5-3. Carnegie Unit Report

School ID: _____ Coder: _____

_____ credits = 1 Carnegie Unit

Explanation:

☐ Explicitly stated in school documents

☐ Inferred from transcript data:
# of credits received for a full year course taken everyday, 1 period.

or

# of credits received for a semester-long course taken every day, 1 period **times 2**

☐ Telephone conference verification

☐ other [explain]:

Date Sources Used: Date: _____

Catalog	☐	Call to School (attach report)	☐
Transcripts	☐	Other:	☐

Any changes over past four years? YES ☐ NO ☐

If yes:

1989 _____ credits = 1 Carnegie Unit

1988 _____ credits = 1 Carnegie Unit

1987 _____ credits = 1 Carnegie Unit

Exhibit 5-4. Standardization of grades

STANDARDIZATION OF GRADES

SCHOOL ID# _____ INITIALS _____

STANDARD	LIST ALL SCHOOL EQUIVALENTS
01 = A+	
02 = A	
03 = A-	
04 = B+	
05 = B	
06 = B -	
07 = C+	
08 = C	
09 = C-	
10 = D+	
11 = D	
12 = D-	
13 = F	
14 = PASS OR SATISFACTORY	
15 = UNSATISFACTORY	
16 = WITHDREW	
17 = INCOMPLETE	
18 = NON GRADED	
19 = BLANK	
OTHERS (Specify)	

NOTE: ATTACH SAMPLE TRANSCRIPT GRADES FOR TRANSFER AND LIST ID NUMBERS. IF APPLICABLE.

5.6.1 Quality Control for Transcript Data Entry

Measures to maintain the quality of data entry on transcripts included (1) 100 percent verification of data entry; (2) review of all transcripts where the number of credits reported for a given year (or the total number of credits) was not indicative of the school's normal course load or graduation requirements; and (3) reconciliation of IDs of transcripts entered with the list of valid IDs for the HSTS. Verification included all data entry fields except for course titles, test names, and award titles. Verification was performed by a CADEr who had not entered that data initially. The number of credits entered for a transcript was automatically compared to a file containing the number of credits required for graduation, and gave the CADEr a warning message if the number of credits entered was too large or small to be feasible. By reconciling the IDs on the transcripts that were entered with the IDs of students on the HSTS-eligible list, we ascertained that every eligible transcript was entered and that no ineligible transcripts were entered.

5.6.2 Quality Control for Catalog Data Entry

The full listing of catalog titles was reviewed by a curriculum specialist who visually compared the listing with the catalog itself. When errors were encountered, corrections were keyed and the corrections were reviewed again. For those schools without catalogs, the listing that was generated automatically was reviewed and edited when courses were coded.

5.6.3 Quality Control for Catalog Coding

Our procedures for assuring the quality of assigning CSSC codes to courses offered in HSTS schools included (1) careful training and supervision of coders; (2) formal reporting and resolution of coding difficulties; (3) reliability checking throughout the process through independent coding of a sample of courses, or by complete review of codes for non-transfer courses by the curriculum specialist; (4) extensive quality reviews; and (5) automated quality assurance reports. Each of these procedures is described separately below. Figure 5-1 is a schematic diagram of our quality control procedures for catalog coding.

5.6.3.1 Personnel Selection, Training, and Supervision

We used trained, experienced educators for the coding task to enable coding to be performed in a meaningful rather than rote manner. These coders had sufficient experience to understand, for example, the subtle differences in levels of English courses (regardless of specific terms used to describe them) so that they would be coded appropriately as at, above, or below grade level, and to recognize what the term "grade level" really means. After selecting individuals with appropriate experience and background, we conducted thorough training (see Section 5.3.3), in the concepts and procedures to be used in performing the coding task. The training included multiple measures of trainees' understanding and accurate use of the information presented. Two of the coders had served in a similar capacity for the 1990 HSTS.

A curriculum specialist, holding a doctorate in Curriculum and Instruction, supervised the entire coding operation. She was constantly available to coders to answer questions, verify information, discuss issues, and provide general guidance as questions and problems were encountered. All issues that were of a general nature (i.e., pertaining to coding many or all catalogs) were brought to the attention of the entire group of coders. Answers to difficult coding decisions were posted on a wall visible to all coders. The curriculum specialist periodically reviewed each coder's work to ensure a continued high level of performance.

5.6.3.2 Difficulty Reporting

A Catalog Coding Difficulty Report (Exhibit 5-5) was sent to the curriculum specialist for review and final resolution whenever a catalog coder encountered a problem. These reports were filled out for all problems, even if they were solved "on-the-spot," to document any difficulties that arose and the decisions that were made. The curriculum specialist annotated the report when the problem

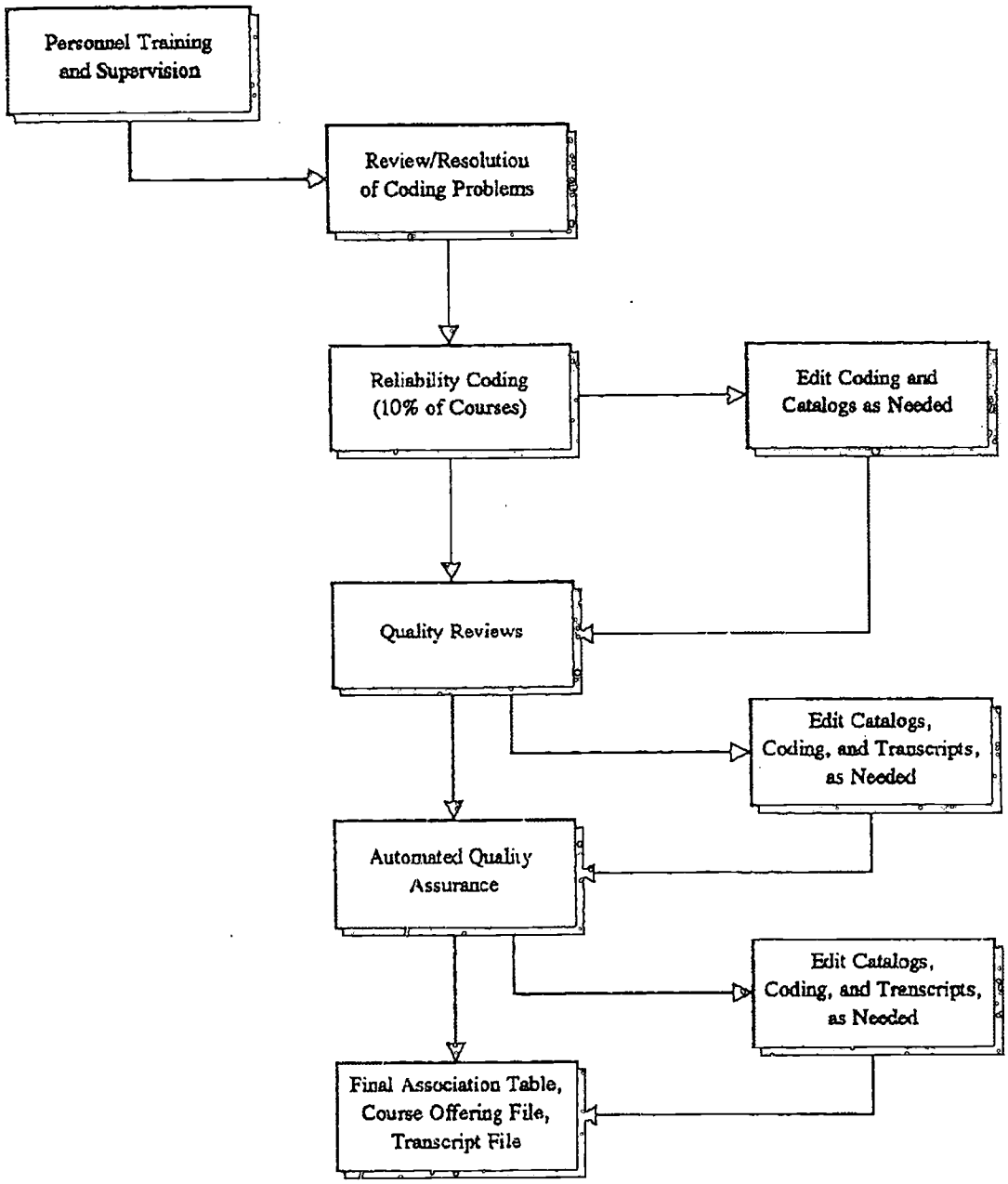


Figure 5-1. Quality control processes for catalog coding

Exhibit 5-5. Catalog coding difficulty report

School ID: _____ Coder: _____

Date: _____ Referred to: _____

Nature of difficulty:

Response:

Date of response: _____ Initials: _____

was resolved, indicating what decision was made. Additional reports of occasional telephone conferences with school personnel were completed, whenever such calls were necessary to answer important questions. Exhibit 5-6 is the form used to document these telephone conferences.

5.6.3.3 Coding Reliability

An important measure of the quality of catalog coding is reliability, or agreement between coders on an appropriate CSSC code for a course. To measure coding reliability, one of the experienced coders coded a random sample of 10 percent of the non-transfer courses in each school catalog.

For schools with fewer than 100 non-transfer titles in their catalogs, 10 courses were coded by the experienced coder. For schools with more than 250 titles, 25 courses were coded. We then compared this sample coding with the codes assigned to the same course by the catalog coder. An agreement is either an exact match of codes or a match to a code that the curriculum specialist determines is equally appropriate for the course. If 90 percent or more of the coding agreed, no further action was taken. If agreement was less than 90 percent, the catalog coding was completely reviewed and any necessary changes were made. The disagreements were also discussed with the catalog coder who had done the original coding, and all coding procedures and principles were reviewed, as necessary. In addition, for 90 percent of the schools, the curriculum specialist reviewed all coding of non-transfer courses and made changes as needed. The coding supervisor filled out a report on reliability coding for each school. Agreement of 90 percent or better was found for approximately 85 percent of the school catalogs during the first review. Since nearly all catalogs were completely reviewed by the coding supervisor and corrected, we ensured that coding accuracy was high. Exhibit 5-7 is a sample of the form used to document coding reliability.

5.6.3.4 Quality Review

Additional procedures to measure and maintain quality included a two-step review process. The first step consisted of generating a report for each school listing the courses that were uncoded, coded as "uncodeable" or coded with an "other" code. Another report listed transcript titles that were unmatched or matched to an "uncodeable" course. The curriculum specialist reviewed all these and re-coded and

Exhibit 5-6. Telephone conference report

School ID: _____ _____ Coder: _____

Phone Number: _____ Date: _____

Contact: _____ Position: _____

Purpose(s) of Contact:

School's Response(s):

Exhibit 5-7. Catalog coding discrepancy report

Coder: _____ School ID: _____ Date: _____

% Agreements: _____ % Diagreements: _____

Matches

Catalog Title	CSSC Title	CSSC Code	Digit 7	Flags

Discrepancies

Catalog Title	Codes Flags	Verified Code	CSSC Title	Error, Match or Flag

Recoding: Coder: _____ Date: _____

re-matched to the fullest extent possible all courses for which she could provide more explicit coding. The second step, or "final review" was the last step in verifying the accuracy and completeness of all coding. The curriculum specialist performed this review by examining each CACE file a final time, paying close attention to title matching, as well as to catalog coding. When this review identified problems, the file was returned to a catalog coder to fix the problems, and the quality review procedures were repeated.

5.6.3.5 Automated Checks

An additional quality check took place when the CACE files for a school were converted to delivery format. Reports listing frequencies of occurrences that might indicate errors were sent to the curriculum specialist to review carefully. Each file was then assigned a status of (1) complete, (2) errors in transcript entry, (3) errors in catalog coding and associations, or (4) computer errors (such as duplicate course sequence numbers). A file with status of 2, 3, or 4 was returned to CADE and CACE for correction, a new report was generated, and the report was again reviewed. This process was repeated until the file had a status of 1, indicating that it was complete and correct.

We reviewed the transcripts and data files of all students with less than 75 percent or more than 150 percent of their schools' graduation requirements to ensure that no entry errors were made. During the review, we found results as described in the remainder of this section.

In a small number of cases, we discovered that a student had not actually graduated and changed his or her exit status accordingly. In another group of cases, we found that some students actually had earned substantially more credits than are required to graduate. Often these were students who had spent substantial amounts of time in both Mexican and American high schools. While they were awarded credit for the Mexican courses, they were still required to take an essentially American curriculum in order to obtain the American diploma.

In still other cases, we found that, although a graduate had fewer credits than were required to graduate, the transcript had all the other attributes of a graduated senior such as 4 full years of courses, all required courses, a graduation date, a grade point average, and a class standing. In these cases, if a careful review of the transcript and the data files showed no data entry or coding errors, we kept the transcript in the database with the apparent inconsistency as recorded on the transcript.

In a small number of cases the transcript being reviewed listed transfer courses that needed special treatment. In some cases it was clear that the appropriate conversion factor for the credits reported on the transcript to Carnegie units was different from that of the school issuing the transcript. When this occurred, we adjusted the conversion factor appropriately for these courses on a student-by-student basis. In other cases, we found entries on transcripts indicating that a student had been awarded some number of credits for transferred courses, but no list of the specific courses. When this happened, we created a dummy course titled "Undifferentiated Transfer Courses" and treated it as an uncodable course.[14]

Inclusion of the Undifferentiated Transfer Courses on the file had the effect of accounting for all the credits that appear on the transcripts. It also provided us with a means of screening essentially incomplete transcripts out of the analyses. Because the intent of the transcript study is to summarize the course-taking patterns of graduates of American high schools over the 3 or 4 years that they are in a typical high school, for analytic purposes we treated transcripts that did not list separate credits for the equivalent of at least three full years of high school courses as incomplete. We did this by creating a flag (GRREQFLG), which we placed on the student file, that indicated whether the differentiated course credits on a transcript totaled at least 75 percent of the minimum credits required to graduate. If they did not, the transcript remained in the file, but the student was given a weight of zero and treated as missing for purposes of projecting national totals (see Section 6.5 for a description of the nonresponse adjustment procedures).

We reviewed all SS transcripts of students with special education diplomas or certificates of attendance with GRREQFLG=4. We determined that 29 of these students had transcripts that listed either three or four years of their high school course work. This situation can occur when a student has an Individualized Education Program. Although these 29 students had unusual graduation requirements, their transcripts represented a portion of the American high school experience. For this reason, we assigned positive final weights to all 29 of them despite the fact that they had fewer credits than other graduates in their schools. These students were, however, treated as ineligible in the computation of student nonresponse and post-stratification adjustment factors. We fully coded the transcripts for such students and provided their data on the file.

[14] If a list of transfer courses appeared on a transcript with a number of credits indicated for the group of courses, catalog coder apportioned the credits among the courses using whatever information was available. For example, some transcripts had sections that indicated by a series of check marks which of a set of requirements were met. If the courses explicitly detailed on the transcript did not account for all of the check marks, then the transferred credits must account for the remainder.

Some of the automated checks performed on the files produced by the transcript data entry and coding process included the following:

- All files were checked for duplicate IDs.
- We verified that all NAEP IDs in the control list also appeared on the TRF list.
- We verified that all IDs on the TRF list for a school were in the student data file.
- We created a crosstabulation of graduation year by exit status and reviewed all outliers.
- We created a crosstabulation of highest year (e.g., 11th grade, 12th grade) appearing in the transcript by exit status and reviewed all outliers.
- We created a crosstabulation of total Carnegie Units earned by exit status and checked all outliers.
- We listed all students with 12th grade transfer courses (other than summer school) and checked their transcripts for accuracy of data entry.
- We checked for valid combinations of course flags. For instance, no course could be both honors and remedial or special education.

5.7 Scanning and Preparing the IEP/LEP Questionnaires

Identical IEP/LEP Questionnaires were used for NAEP and HSTS, and most of the questionnaire items needed no recoding. The responses were entered on optical scan forms by school personnel (see Section 4.5) and scanned by NCS. The data in the scanned data file were direct representations of the questionnaire responses. There were, however, four items on the scanned data file that needed some recoding. The same recoding algorithm was used for the following three items:

Item 4. What percentage of the school day does this student spend in a regular class?

Item 7. What percentage of the school day is this student served by a special education program?

Item 18. What percentage of the school day is this student served by a special language program?

The choices on the questionnaire were 0 percent, 10 percent, 20 percent, and so on through 90 percent and 100 percent. For each item, the scanned data file contained one variable (coded "Yes" or "Missing") for each possible percentage choice. Because of this, it was possible to have more than one percentage entered in response to Questions 4, 7, and 8. The following actions were taken in order to create a file with a single field containing the actual percentage indicated on the questionnaire.

- If the respondent checked a single response for the item, the value of that response was used;
- If the respondent checked two adjacent responses, they were averaged;
- If the respondent checked more than two responses or two non-adjacent responses, the response code for "multiple response" was used; and
- If no response was checked, the code for "missing" was used.

We also recoded one other item from the scanned data file:

Item 6. Which of the following best describes this student's disability?

Once again, the scanned file is structured in such a way that each possible selection is a separate variable. This allowed multiple selections to occur. Our solution was to recode the responses in the following manner:

- If the respondent checked multiple responses and they were "visually HC/blind" and "deaf/blind," then the response became "deaf/blind" and
- In any other case where two or more responses were chosen, the code for "multidisabled" was used.

Several variables were added to the final IEP/LEP file. The student disability status was determined by the students' IEP status as indicated by the first question on the questionnaire and by the pattern of answers to the content questions. The disability flag (HCFLAG) was set to "1" if no disabling condition was indicated in our records, otherwise it was set to "2". Specifically, the disability flag was set to "2" if the following conditions were met:

- The TRF had the IEP field flagged as 1 ("Yes");
- The student's exit status as entered in the CADE system is 3 or 4 (special education diploma or certificate of attendance);

- Question 1 ("Why is this student classified as IEP/LEP?") in the IEP/LEP Questionnaire had response 1 ("A. A disability (physical or mental disability)") or 3 ("C. Both a disability and limited English proficiency"); or
- Question 1 in the IEP/LEP Questionnaire is not 1 or 3, but a specific disabling condition identified in Question 5 and Question 7 indicated that the student was being served by a special education program for some portion of the day.

The students' exit status, race/ethnicity, grade level, sex, birth month and year, and Chapter 1 flag were obtained from the Student File. If that information did not exist on the Student File, the corresponding data from the IEP/LEP questionnaire were incorporated if available. Frequencies and crosstabulations were run to check the data for valid entries and outliers before, during, and after processing. For the data collected specifically for the HSTS, unusual values were rechecked against the original documents and corrected as necessary.

5.8 Scanning and Preparing the School Characteristics and Policy Questionnaires

The School Characteristics and Policy Questionnaire (SCPQ) was used in the 1994 NAEP and was available for 282 of the 340 HSTS schools (the remainder had either not participated in NAEP or had failed to respond to the questionnaire). An additional 43 SCPQs were gathered by Westat during the transcript data collection. Fifteen schools did not complete SCPQs. The data were entered on optical scan forms by school personnel and scanned by NCS.

When coding the SCPQs, the coding system used in the 1987 and 1990 School Files was used whenever possible. As with the IEP/LEP Questionnaire, processing consisted of reformatting the scanned responses to provide one variable per question. When necessary, the value was set to either "multiple response" or "no response" as appropriate.

A copy of the 1994 SCPQ is included as Appendix B. The *1994 High School Transcript Study Data File User's Guide* provides a complete list of the variables on the SCPQ and their values. This information has been incorporated into the School File.

6. WEIGHTING AND ESTIMATION OF SAMPLING VARIANCE

The 1994 High School Transcript Study used a complex sample design with the goal of securing a sample from which estimates of population and subpopulation characteristics could be obtained with reasonably high precision (in other words, low sampling variability). At the same time, it was necessary that the sample be economically and operationally feasible to obtain. The resulting complex sample design requires that the user of the HSTS data utilize sampling weights to ensure valid analysis of the transcript data.

Sampling weights are factors assigned to each transcript which are used in any aggregations of transcript characteristics. Heuristically, these weights can be seen as being the number of students in the population that the sampled transcript "represents." A transcript with a sampling weight of 100 represents 1.0 the sampled student and 99 other nonsampled (or sampled but nonresponding) students in the population. A transcript with a sampling weight of 1 represents only the sampled student.

The sampling weights are designed primarily to represent differential sampling and response rates. For example, if a student comes from a subcategory with a sampling rate of 1/10 and a response rate of 1/2, then the student's transcript might receive a sampling weight of 20. That transcript can be seen as representing the student and 19 other nonsampled and nonresponding students.

From the viewpoint of assigning sampling weights, the most important aspect of the 1994 HSTS sample design was the utilization of differential sampling rates. For example, schools with high percentages of minority students were sampled at a doubled sampling rate, and very small schools were sampled at a lower rate to reduce the costs incurred in fielding the schools (see Chapter 2 for further details regarding the sample design). Section 6.1 discusses the procedure for assigning sampling weights.

One consequence of the HSTS sample design is its effect on the estimation of sampling variability. Because of the effects of multistage design (students within schools, schools within primary sampling units) and because of the effects of certain adjustments to the sampling weights (poststratification and weighting adjustments), observations made on different students cannot be

assumed to be independent of one another. As a result, ordinary formulas used to estimate the variance of sample statistics, based on assumptions of independence, will tend to underestimate the true sample variability. Three techniques which are widely utilized for variance estimation under those circumstances are linearization, balanced repeated replication (BRR), and the jackknife. The jackknife procedure provides reliable variance estimators while being easy for the user to utilize. Any aggregations are computed utilizing the original sampling weights and each set of jackknife replicate weights. A simple formula combines these estimates into a suitable variance estimator.

Two types of weights, HSTS sample weights and linked weights, are needed for these data. HSTS sample weights are designed for any aggregations, including all of the transcripts in the study, whether or not they correspond to assessed NAEP students. The weight of each transcript represents students not included in the HSTS Study. Linked weights are designed for any aggregations which only include transcripts from students who were in a particular NAEP assessment (or who were excluded from NAEP). In this case, the linked weight assigned to the transcript is designed to represent not only students not included in the HSTS study, but also students included in the HSTS study who were not given the same assessment.

6.1 The HSTS Sample Weights: An Introduction

In order to make valid inferences about the entire population of graduated grade 12 students from the sample of student transcripts collected, it is necessary to use the sampling weights. The weights reflect the probability sampling scheme used to arrive at the sample of students for whom transcripts were requested. The weights also reflect the impact of sample nonresponse at the school and the student level, and make adjustments for these groups to decrease the potential bias that might arise through differential nonresponse across population subgroups. Finally, improvements to the precision of weighted estimates result from the application of poststratification factors to the sample weights.

Since the derivation of sampling weights and the estimation of sampling variability are strongly related to the sample design, the reader will need to review the main features of the sampling design discussed in Chapters 2 and 3 of this report.

The final HSTS student weight was constructed in four steps. The first step was to construct the student base weight (or design unbiased weight), which is the reciprocal of the overall probability of selection. This procedure is discussed in Sections 6.4.1 and 6.4.2.

The second step was to compute school nonresponse factors, adjusting for schools that did not participate in the HSTS study. This procedure is discussed in Section 6.5.

The third step was poststratification. Poststratification is the process of adjusting weights proportionally so that they aggregate within certain subpopulations to independent estimates of these subpopulation totals. These independent estimates were obtained from the Current Population Survey (CPS) estimates for various student subgroups. For example, one poststratification subcategory was Hispanic students. The CPS estimate of the number of Hispanic students is 159,200. The corresponding aggregation of the sampling weights is 144,800. The sampling weights for Hispanics are all adjusted by the factor 159.2/144.8 so that the sampling weight aggregation also equals 159,200. As the CPS estimate has smaller sampling error associated with it, this adjustment should improve the quality of the weights. This step is discussed in Section 6.6.

The final step was to adjust the poststratification student weight for the graduated students with transcripts to account for students with missing transcripts. This process is discussed in Section 6.7.

The linked student weights were constructed in a parallel manner, with some differences. For example, the student base weight incorporated a factor for assignment to NAEP assessments (discussed in Section 6.4.3).

The school nonresponse factors were also slightly different than the corresponding HSTS student weight school nonresponse factors, to account for schools that refused to participate in NAEP. Section 6.5.5 presents a discussion of school nonresponse factors.

There was an extra nonresponse factor computed for the linked weights not included in the HSTS weighting computation. This was an adjustment for students whose transcripts were included in the HSTS study, but who were absent from, or refused to participate in, a NAEP assessment. This adjustment is discussed in Section 6.6.1.

The trimming and poststratification steps for the linked weights were similar to those of the HSTS weights, with some differences. These steps for the linked weights are discussed in Sections 6.6.3, 6.6.5, and 6.6.6.

Finally, the missing transcript adjustments for the linked weights were very similar to those computed for the HSTS weights. These are discussed in Section 6.7.2.

6.2 Variance Estimation

For variance estimation, both the 1994 NAEP survey and the 1994 HSTS survey used the jackknife technique which, as its first step, draws carefully selected subsets of the data. For each respondent in each subset a sampling weight is determined, as if the chosen subset were in fact the responding sample. The recomputation is complete, including a generation of new nonresponse adjustments and new poststratification adjustments using only the subset. This process generates a set of "replicate" weights for each responding sample member. These replicate weights are used to compute a series of replicate estimators for each survey characteristic. The variability of these replicate estimators around the original estimator gives a reliable measure of the sampling variance of the original estimator.

A considerable amount of theoretical and empirical work justifies the jackknife technique as a variance estimation method for surveys such as the 1994 HSTS survey. In cases where the variance estimator is simple, the jackknife estimator is usually equal to this variance estimator. Thus, in this situation, the jackknife would be redundant. The jackknife is valuable because it is also reliable as a variance estimator when the "correct" variance cannot be computed at all, as is the case with the 1994 HSTS survey. There is a wide range of literature discussing the jackknife; a good general overview of the theory is given in Wolter (1985), Chapter 4.

The jackknife procedure is generally used at Westat for surveys such as the 1994 HSTS survey. Westat has used this method for calculating sampling errors for a wide range of survey designs. Besides being known to be generally reliable, it is relatively straightforward for secondary analysts to calculate sampling errors appropriately. For any given survey characteristic, an analyst would need only to generate a series of estimators using the replicate weights and the original weights. The variance estimator would then be computed using these "replicate estimators." In particular, the

analyst does not need to have a complete understanding of the sample design and weighting procedures to calculate these variance estimators accurately.

The 1994 NAEP survey used 62 replicate weights for computation of jackknife variance estimates. As already noted, the 1994 HSTS sample was a subsample of the schools selected into the 1994 NAEP sample. The replicate weights were generated by randomly deleting sampling units at the first stage of sampling. The sampling weights were then recomputed without these randomly deleted replicate groups. For the noncertainty PSUs, the first stage of sampling was at the PSU level, requiring that the deleted units be sampled PSUs. Thirty-six of the NAEP replicate weights were generated by deleting one sampled PSU from a pair of sampled noncertainty PSUs. Since the HSTS is based on the same sample of noncertainty PSUs, HSTS replicate weights are based on the same set of replicate groups.

There was one noncertainty PSU that had no sampled HSTS schools. The PSU that was paired with this school for variance estimation purposes in NAEP was re-assigned to another pair (making the pair a triplet). The HSTS survey therefore has only 61 replicate weights rather than 62, with 35 associated with noncertainty PSUs. See Section 6.4.5 for more details.

A different situation existed for the certainty PSUs. For those, the first stage of sampling was at the school level: the deleted units were sampled schools rather than sampled PSUs. Twenty-six of the NAEP replicate weights were generated by deleting a set of sampled schools from the set of sampled schools in the certainty PSUs. Since the HSTS sample of schools was a random subsample taken from the original NAEP sample of schools, we created HSTS replicate groups by deleting random groupings of the HSTS schools in each certainty PSU. This approach gave us 26 of the 61 replicate weights for the 1994 HSTS study.

The Degrees of Freedom of the Variance Estimate

It is important to have an indication of the number of degrees of freedom to attribute to the jackknife variance estimator v(t) of Var(t). The degrees of freedom of a variance estimator provide information on the stability of that estimator: the higher the number of degrees of freedom, the lower the variability of the estimator. In practical terms, the number of degrees of freedom of the variance

estimator corresponds to the number of residual degrees of freedom that can be assumed for inferential procedures.

Since the jackknife procedure estimates the sampling variability of the statistic by assessing the effect of change in the sample at the paired first-stage sampling unit (FSSU) level, the number of degrees of freedom of the variance estimator v(t) is at most equal to M, the number of FSSU pairs. The maximum number of degrees of freedom equals the number of independent pieces of information used to generate the variance. In the case of data from the main assessments, the pieces of information are 62 squared differences $(t_i - t)^2$, each supplying at most one degree of freedom (regardless of how many individuals were sampled within any FSSU).

The number of degrees of freedom of the sample variance estimator can be strictly less than the number of FSSU pairs. For example, suppose that the statistic t is a mean for some subgroup, and no members of that subgroup can come from either FSSU in the i^{th} FSSU pair. (Examples of such subgroups are any PSU-level partitioning of the population, such as region.) In this instance, neither member of the FSSU pair i directly contributes to the estimate of t, so that the pseudoreplicate t_i would nearly equal the statistic t. If the replicate weights used to generate t_i had not received poststratification adjustments, the resulting pseudoreplicate t_i would be identical to the overall estimate t so that $(t_i - t)^2 = 0$. In this case, such an FSSU pair would impart no information on the variability of the statistic t and thus contribute 0 degrees of freedom to the variance.

Our approach regarding the 1994 HSTS survey is to err on the side of being overly conservative in assigning degrees of freedom. For any estimate of the full population, we recommend using confidence intervals based on the t distribution with 25 degrees of freedom. This is probably conservative, but there is little practical difference between confidence bounds for t distributions with more than 25 degrees of freedom.

For estimates of subpopulations that are national (not concentrated in a single region), we recommend confidence intervals based on the t distribution with 10 degrees of freedom. Again this is likely to be conservative for most subpopulations based on gender, race/ethnic status, urban/rural status, and so forth, which are represented within most of the FSSU pairs in the study.

6.3 The HSTS-NAEP Linked Weights: An Introduction

A primary purpose of the HSTS study is to provide a database for analyzing the relationship between students' proficiencies, as measured by their NAEP assessment outcomes, and students' course-taking in their high school careers. In order for a student to be part of this "linked" database we required a completed NAEP assessment for the student, as well as a completed transcript from the HSTS study. There were many students for whom we have a completed transcript, but no NAEP assessment (due to a refusal of either the school or the student to participate in NAEP). These students can be part of the HSTS database but not the linked database that requires both transcripts and assessment results for the same student.

The linked database requires a different set of sampling weights than the HSTS database alone, as the set of students that qualify tor this database is a subset of the larger HSTS set. In particular, the school and student nonresponse adjustments will be larger for the linked weights than for the HSTS weights. This is so because a student or school had to participate in both the NAEP and the HSTS surveys to qualify as a "respondent" for the linked data base, reducing the number of both school and student respondents (the nonresponse adjustments are larger when the set of respondents is smaller).

The sampling weights are computed so that the sample can "represent" in a statistical sense the full population of students from which the sample is drawn. In particular, the sampling weights will aggregate to the total number of students in the population. Linked weights are computed separately for reading, history, and geography assessment students. Each assessment sample represents the full population, so each of the three sets of assessment linked weights aggregate separately to the population totals. A separate set of linked weights is also computed for excluded students. The summation of these weights over all excluded students in the sample is an estimator of the total number of students in the population who would have been excluded from the NAEP assessment if the full population had been included in the study (rather than a sample).

6.4 Computation of the Base Weights

Sample estimates were computed from the students' transcripts by aggregating observations from each transcript using the sample weights. If there were 100 percent response to the

HSTS survey, and if no poststratification were carried out, then the sample weights would be equal to the base weights, which are the reciprocals of the probabilities of selection of that student. The sample aggregates generated using these base weights would be unbiased estimators of the corresponding quantities in the U. S. population (cite, for example, Cochran (1977), Section 9A.7).

6.4.1 Computation of Base Weights: HSTS Weights

The student base weight for the 1994 HSTS sample was computed for each student sampled into a NAEP assessment (including selected students who were later excluded as being nonassessable), in an HSTS sample school. The weight was computed as the reciprocal of the overall probability of selecting the k-th student from the j-th school and i-th PSU, which is the product of three weights:

$$w_{ijk} = w_i w_{j|i} w_{k|ij}$$

where,

$$w_i = 1/p_i, \quad w_{j|i} = 1/p_{j|i}, \quad w_{k|ij} = 1/p_{k|ij},$$

p_i is the probability of selection of the i^{th} PSU, (see Section 2.2)

$p_{j|i}$ is the conditional probability of selection of the j^{th} school into the HSTS sample, given that the i^{th} PSU was sampled,

$p_{k|ij}$ is the conditional probability that student k was sampled within school j in PSU i.

$p_{j|i}$ has two factors: the conditional probability of selection of the school into the 1994 NAEP sample, given that the sample PSU was selected (see Section 2.2), and the conditional probability of selection of the school being selected into the HSTS sample. The 'frame' for the HSTS sample was the set of all eligible 1994 NAEP sample schools which were sampled for the primary NAEP Age 17 Study. The HSTS sample schools were drawn from this set as a stratified equal probability sample with two strata: public and private schools. The sampling fraction for public schools in this set was .88167, and the sampling fraction for the private schools in this set was .29389.

For schools which participated in NAEP, p_{klj} is the probability the student was sampled to be assessed in NAEP (see Section 2.5).

Table 6-1 presents the following information for public, Catholic, and non-Catholic private schools:

1. The number of schools in the 1994 NAEP main age 17 sample.
2. The number of schools in the first set which were found to be eligible for NAEP.
3. The number of schools in the second set that were sampled into the HSTS sample.
4. The percentage of the third count as a fraction of the second count.

Table 6-1. Counts of NAEP and HSTS sampled schools

School Type	Sampled NAEP schools	Eligible NAEP schools	Sampled HSTS schools	Percentage of eligible NAEP schools sampled
Public	398	379	332	87.6
Catholic	46	45	14	31.1
Non-Catholic	218	114	33	28.9
Total	662	538	379	70.4

6.4.2 Conditional Student Base Weights for the HSTS

As noted before, the quantity p_{klj} is the conditional probability of selection of the student into the NAEP sample for the school, for any schools that participated in the 1994 NAEP assessment. In schools that did not participate in the NAEP assessment, but did participate in HSTS, a sample of students was drawn for the HSTS survey alone. There were 57 of these schools, representing 15 percent of the HSTS sample. If the school had fewer than 60 12th-graders, then the sampling rate was set to 1. Otherwise, an equal probability sample of 50 12th-graders was chosen and the conditional probability of selection was 50 divided by the total count of 12th-graders in the school.

There were also three schools which were cooperative with the NAEP assessment, but did not retain the administrative information necessary to use their assessed students in the HSTS study. New samples of transcripts were taken for these three schools in the same way as was done for the NAEP noncooperating schools.

Table 6-2 presents the total number of students in the HSTS study from each class of school.

Table 6-2. Total students in HSTS study in HSTS cooperating schools

Response Category	Number of schools in category	Number of students in HSTS study
HSTS and NAEP cooperating schools	280	25,904
HSTS cooperating, but not NAEP	57	2,695
HSTS cooperating, no NAEP link	3	216
Total	340	28,815

The schools in the first group are called "linked" schools: students in these schools receive positive sample HSTS and linked weights. Students in the remaining schools receive positive HSTS sample weights. but linked weights of 0.

6.4.3 Computation of Base Weights: NAEP-HSTS Linked Weights

The student base weights appropriate for the NAEP-HSTS link are similar to those computed for the HSTS weights. However, the probability that a school was assigned the particular session and the probability that a student was assigned to the particular session must also be included as subsampling was done to select final school and student samples for each assessment.

Each student was assigned one of three assessments (to minimize the workload required for each student). This assignment was random. After this assignment, the student was evaluated as to eligibility and excluded from assessment if found to be ineligible (because of language problems or disabilities). Each student was assigned to one of the three assessments, or excluded from any assessment. The sets of students assigned to each assessment are designated U_1, U_2, and U_3,

respectively. The students excluded from any assessment are designated U_e. An indicator function, I, is defined as follows. For any of the four sets (for example, U_1):

$$I[ijk \in U_1] = \begin{cases} 1 & \text{if student } ijk \text{ is in set } U_1. \\ 0 & \text{otherwise.} \end{cases}$$

A base weight can be assigned for each assessment group for each student. This weight is defined as zero (0) if the student was not in that assessment group. The assignment a base weight assigned to student ijk is as follows:

$$w_{ijk}^{a} = w_i w_{j|i} w_{a|ij} w_{k|ij} w_{a|ijk} I[ijk \in r_a]$$

where

$$w_i = 1/p_i, \quad w_{j|i} = 1/p_{j|i}, \quad w_{a|ij} = 1/p_{a|ij}, \quad w_{k|ij} = 1/p_{k|ij}, \quad w_{a|ijk} = 1/p_{a|ijk}$$

p_i is the probability of selection of the i^{th} PSU,

$p_{j|i}$ is the conditional probability of selection of the j^{th} school into the HSTS sample, given that the i^{th} PSU was sampled,

$p_{a|ij}$ is the conditional probability that at least one session of type a was assigned to school j,

$p_{k|ij}$ is the conditional probability that student k was sampled within school j, and

$p_{a|ijk}$ is the conditional probability that student k in school j was assigned to session type a.

Remembering that $w_{ijk} = w_i w_{j|i} w_{k|ij}$, the weight w_{ijk}^{a} can also be written in terms of the HSTS base weight w_{ijk}. See Section 6.4.1 for the definition of w_{ijk}:

$$w_{ijk}^{a} = w_{ijk} w_{a|ij} w_{a|ijk} I[ijk \in U_a]$$

For excluded students, selection into assessment groups is irrelevant. The excluded student base weight can be written as:

$$w^{e}_{ijk} = w_{ijk} I\left[ijk \in U_e\right]$$

In other words, for excluded students, linked base weights are the same as their HSTS base weights. Note that each student in principle is assigned all four weights: the three assessment weights and the excluded student weight. However, for a given student only one of these weights will be nonzero: one of the assessment weights if the student was assessed, or the excluded student weight if the student was excluded.

6.4.4 Conditional Session Probabilities

As discussed in the previous section, the conditional probability p_{alij} is the probability that at least one reading session or at least one history-geography session was assigned to the school. (History and geography assessments were assigned together in joint sessions.) This section briefly presents details regarding these probabilities.

Most schools had sessions of both kinds assigned. For these schools p_{alij} is equal to 1. There were some smaller schools (mostly private) which were assigned only one session (either reading or history/geography). In each of these cases, p_{alij} was equal to 1/2. See Section 2.4 for details regarding session assignments. Table 6-3 presents the counts of schools in each of these groups. This count includes only schools with students with positive linked weights ("linked schools").

Table 6-3. Session statuses for public and private linked schools

Type of school	Both sessions	Reading session only	History/ geography session only	Total linked schools
Public	230	10	6	246
Private	19	7	8	34
All Schools	249	17	14	280

If the school was assigned sessions of both types, then generally a student had a 1/2 chance of being assigned to a reading session and a 1/2 chance of being assigned to a history/geography session. In some of the smaller schools there was an imbalance between the number of reading sessions and the number of history/geography sessions. The probability a student had of being assigned to each session was something other than 1/2 in these cases.

Table 6-4 presents the percentages of students in the HSTS study in the linked schools coming from schools with differing probabilities of students being assigned to a reading session. (The probability for each student being assigned to a history/geography session is 1 minus this reading session probability.) In other words, Table 6-4 presents the percentages of students with varying values of this session assignment probability. (Note that for schools with only a reading session or only a history/geography session, the probability of a student being assigned that session is automatically 1. Also note that in this case p_{alg} is 1/2 for that student.)

Table 6-4. Percentages[1] of linked school students with differing values of the reading assessment probability

Reading Session Probability	Percentage of students
School had history/geography session only	0.7
School had reading session only	0.7
Reading session probability between .625 and .75	3.1
Reading session probability 0.6	1.5
Reading session probability 0.5	90.0
Reading session probability 0.42[2]	2.6
Reading session probability 1/3	1.4

[1] This percentage is of the total set of 25,904 HSTS students in the 280 linked schools.
[2] This includes a small percentage (0.1 of total) with an RSP of 0.389.

The final component of the student's assessment base weight is the assignment of the student to either a history or a geography assessment if he or she was assigned to a history/geography session. This probability is always 4/9 for the geography assessment and 5/9 for the history assessment.

For reading assessment students, the probability p_{aijk} is equal to the reading session probability (the probability that the student was assigned to a reading session). For history and geography session students, the probability p_{aijk} is equal to the product of the probability the student was assigned to a history/geography session and the probability the student was assigned the particular assessment (either 4/9 or 5/9).

Table 6.5 gives the final counts of students assigned each type of assessment. These counts are then separated out into two subcounts: students who were excluded from being assessed based on disability, and students who were certified as eligible for assessment.

Table 6-5. Assessed and excluded students in linked schools

NAEP Assessment	Assessed students	Excluded students	Total students
Reading	12,528	462	12,990
History	6,905	244	7,149
Geography	5,571	194	5,765
All Assessments	25,004	900	25,904

6.4.5 Computation of Replicate Base Weights

As discussed in Section 6.2, 61 replicate weights were generated for variance estimation purposes (one less than 1994 NAEP). This section discusses school, HSTS student, and linked replicate base weights.

The school weights are designated as $w_{ij}(r)$, $r=1,...,26$, $r=28,...,62$. The replicate group corresponding to r=27 is the NAEP noncertainty NAEP PSU pair which was dropped. For $r=1,...,26$, and $r=28,...,36$ these replicate weights correspond to pairs of noncertainty PSUs (see Section 6.2). Write $S(r)$ as the set of sampled noncertainty PSUs (first stage sampling units) corresponding to replicate weight r. Except in the case of replicate weight 26, one of the two PSUs is randomly selected to be random half sample group 1 (both PSUs having equal probability of being selected for deletion). This PSU is indicated as S(r,1), the other PSU as S(r,2). The replicate school base weight for the j-th school in the i-th PSU for the r-th replicate weight is computed as follows:

$$w_{ij}(r)=\begin{cases} w_{ij} & i \notin S(r) \\ 2w_{ij} & i \in S(r,1) \\ 0 & i \in S(r,2) \end{cases} \qquad r=1,\ldots,25,\ r=28,\ldots,36.$$

In the special case of the "triplet" of PSUs corresponding to r=26 one of the PSUs was randomly assigned to random half sample group 1, S(26,1), and one to random half sample group 2, S(26,2). The remaining PSU is designated as S(26,3). The replicate weights assigned for r=26 are then as follows:

$$w_{ij}(26)=\begin{cases} w_{ij} & i \notin S(26) \\ 1.5w_{ij} & i \in S(26,1) \\ 0 & i \in S(26,2) \\ 1.5w_{ij} & i \in S(26,3) \end{cases}$$

For r=37,...,62, the replicate weights correspond to certainty PSUs. The replicate groups for these replicate weights correspond to sets of schools rather than to PSUs, as schools are the first stage sampling units for certainty PSUs (see Section 6.2). Write *S(r)* as the set of schools corresponding to replicate weight r.[15] The replicate groups are generated by randomly assigning one of the half sample groups of schools to random half sample group 1 for pairs, and randomly assigning two of the three groups of schools to random groups 1 and 2 for triplets.[16] (See also Section 6.2). These random half sample groups will be indicated as S(r,1) and S(r,2), with an S(r,3) also for the triplets. After this random selection has taken place the replicate school base weight for the j-th school, i-th PSU, r-th replicate weight, is computed as follows for the pairs:

$$w_{ij}(r)=\begin{cases} w_{ij} & ij \notin S(r) \\ 2w_{ij} & ij \in S(r,1) \\ 0 & ij \in S(r,2) \end{cases} \qquad r=37,\ldots,62,\ r\neq 42,52,54,57,58,59,60,62.$$

For the replicate weights corresponding to triplets (PSUs with three HSTS sample schools), the computation of base weights is more elaborate. Each of these PSUs is assigned a

[15] For all but two of the certainty PSUs, this set corresponds to all schools in the PSU. For the Los Angeles and New York PSUs, this set for each replicate weight corresponds to a third of the schools in the PSU.

[16] As discussed in Section 6.2, these "triplet" certainty PSUs had three HSTS schools. These PSUs correspond to replicate weights 42, 52, 54, 57, 58, 59, 60, and 62: see Table A-6.2.2 in the Appendix.

companion PSU as follows: 42 and 52, 54 and 62, 57 and 58, and 59 and 60. The assignment of replicate weights is described for replicate weights 42 and 52; the procedure is identical for the other three pairs.

$$w_{ij}(42) = \begin{cases} w_{ij} & i \notin S(42), S(52) \\ 1.5w_{ij} & i \in S(42,1) \\ 0 & i \in S(42,2) \\ 1.5w_{ij} & i \in S(42,3) \end{cases} \quad \begin{matrix} 1.5w_{ij} & i \in S(52,1) \\ 1.5w_{ij} & i \in S(52,2) \\ 0 & i \in S(52,3) \end{matrix}$$

$$w_{ij}(52) = \begin{cases} w_{ij} & i \notin S(42), S(52) \\ 1.5w_{ij} & i \in S(52,1) \\ 0 & i \in S(52,2) \\ 1.5w_{ij} & i \in S(52,3) \end{cases} \quad \begin{matrix} 1.5w_{ij} & i \in S(42,1) \\ 1.5w_{ij} & i \in S(42,2) \\ 0 & i \in S(42,3) \end{matrix}$$

The HSTS student weights and linked weights can now be computed as discussed in Section 6.3. The replicate weight w_{ijr} replaces w_{ij} in these computations.

6.5 Weighting Adjustments for School Nonresponse

Nonresponse is present to some degree in every large-scale survey. This generally has a negative effect on the quality of estimators, if not adjusted for in the weights. First of all, nonresponse reduces the effective sample size from n to n_r, where $n_r < n$. This reduction of sample size increases the sampling variance of any estimators. In addition, if there are significant differences between the respondents and nonrespondents, then there will also be a bias of unknown size and direction. For example, suppose that the overall response rate was 60 percent, but the response rate of black students was only 20 percent, whereas the response rate of white students was 80 percent. Without any adjustment, whites would be overrepresented in the data set by a factor of 4. If there are systematic differences between whites and blacks with regard to any of their HSTS characteristics, then this overrepresentation would result in serious bias. In this example, a nonresponse adjustment would correct this bias by multiplying the sampling weights for black students by a factor of 4.

Suppose Y is the population characteristic of interest, and is the summation of the characteristic value for each student over all graduates in the U.S. population. One such characteristic,

for example, would be whether the student has taken Advanced Placement Calculus. If y_{ijk} is the characteristic value (equal to 1 if the student has the characteristic, 0 otherwise) for the k^{th} student in the j^{th} school in the i^{th} PSU, with P the set of all schools in the U.S. population (in all PSUs), and P_{ij} the set of all graduates in the j^{th} school in the i^{th} PSU, then we can write Y as:

$$Y = \sum_{ij \in P} \sum_{k \in P_{ij}} y_{ijk} \qquad \text{(Equation 6.5.1)}$$

Suppose S is the HSTS sample of schools, with S_{ij} the set of all sampled students in HSTS school j in PSU i. Then under full response we can write the unbiased estimator of Y as:

$$\hat{Y}_F = \sum_{ij \in S} \sum_{k \in S_{ij}} w_{ijk} y_{ijk} \qquad \text{(Equation 6.5.2)}$$

where w_{ijk} is the student base weight for sampled student k in HSTS school j in PSU i. (See Section 6.4 for the definition of w_{ijk}.)

In the HSTS survey there was nonresponse at both the school and the student level. Let RS be the set of cooperative HSTS schools, and RS_{ij} the set of sampled students for which we have completed transcripts in school ij (the j^{th} school in the i^{th} PSU). Then our final estimator of Y can be written as:

$$\hat{Y} = \sum_{ij \in RS} \sum_{k \in RS_{ij}} W_{ijk} y_{ijk} \qquad \text{(Equation 6.5.3)}$$

The weight W_{ijk} in Equation 6.5.3 is the final sampling weight: the base weight w_{ijk} multiplied to adjustments for school nonresponse and missing transcripts at the student level. W_{ijk} also includes factors incorporating poststratification adjustments. The final adjustments for missing transcripts at the student level are discussed in Section 6.7, and the poststratification adjustments are discussed in Section 6.6. The remainder of Section 6.5 discusses the adjustments made in the base weights to account for school nonresponse. It is divided into the following sections:

- Approach to school nonresponse adjustments;
- Selection of school nonresponse cells;

- The results of the CHAID analysis;[17]
- HSTS school nonresponse adjustments; and
- School nonresponse adjustments for the NAEP-HSTS linked weights.

6.5.1 Approach to School Nonresponse Weighting Adjustments

The most widely accepted paradigm for nonresponse weighting adjustments is the quasi-randomization approach (Oh and Scheuren (1983)). In this approach, nonresponse cells are defined based on characteristics of the schools that are known to be related to response. For example, if it is known that private schools generally respond at a lower rate than public schools, then public/private status should be one characteristic used in generating nonresponse cells.

Under this approach, all schools in the sample are assigned to a nonresponse cell c based on their characteristics. The weighting adjustment for each cooperative school will be equal to W_c / W_{rc} , where W_c is a weighted count of graduates in HSTS schools in nonresponse cell c, and W_{rc} is a weighted count of graduates in the cooperative HSTS schools in the same cell. This weighting adjustment is the reciprocal of a weighted response rate of the HSTS school's response cell.

Under the quasi-randomization paradigm, we model nonresponse as if it were equivalent to another stage of sampling. Within each nonresponse cell we assume that the responding schools are a simple random sample from the set of all HSTS schools in the cell. In other words, there are no systematic differences in nonresponse rates within subcategories contained in each cell. If this assumption is valid, then the use of the quasi-randomization weighting adjustment will eliminate any nonresponse bias.[18]

The critical assumption under this approach is that the response rate is homogeneous within the nonresponse cells. For example, if the nonresponse cells are based only on public/private school status, and there are considerable differences in response rates between high minority and low minority schools, then this divergence of response rates within the public/private cells will cause bias in the study results. On the other hand, we only want nonresponse cells for which the response rate is

---

[17] See Section 6.5.2 for a description of CHAID.

[18] (For further discussion regarding these assumptions and model see Little and Rubin (1987), Section 4.4.

in fact heterogeneous across cells. Using more cells rather than less could increase variability and, if many of the cells have the same underlying response rate, then no bias reduction will be achieved by having the larger number of cells. Therefore, we will choose nonresponse cells that are homogeneous in response rate within cells and heterogeneous between cells. We will also choose a set of cells that is as small in number as possible while satisfying these properties.

6.5.2 Selection of School Nonresponse Cells

All eligible responding schools within each selected nonresponse cell receive the same school nonresponse weighting adjustment to their weights. This nonresponse adjustment is formally defined in Section 6.5.4, Equation 6.5.5. It is important that response rates be as uniform as possible within each nonresponse cell. For example, suppose that the nonresponse cells are based on Census region alone, so that Northeast Census region would be one nonresponse cell. Then all schools within the Northeast region would receive the same school nonresponse weighting adjustment, say 1.5. This nonresponse adjustment would be the reciprocal of a response rate of 2/3.

However, suppose that high minority schools within this cell have a response rate of 1/5, with low minority schools having a much higher response rate of 9/10. Then low minority schools would be overrepresented in this sample by a factor of 9/2, and a nonresponse bias would be incurred for any characteristic that is related to minority status. The response rate is not uniform within the response cell, but may be uniform within response cells defined by both Census region and minority status. In this case, the small number of high minority schools would receive a school nonresponse adjustment of 5, with the large number of low minority schools receiving a school nonresponse adjustment of 1.11. High and low minority schools would then be represented correctly in the final estimators.

This need for a uniform response rate within cells requires us to make nonresponse cells as small as possible to capture every characteristic that may be related to both 'response propensity' and survey characteristics of interest. However, at the same time, it is important that the sample sizes within individual response cells do not become too small, because this could seriously increase sampling variability. Thus, we need to assign nonresponse cells that are homogeneous in response propensity within cells, but also have reasonably large sample sizes within each cell.

There are five potential nonresponse variables (for schools and PSUs) that we checked in our analysis.

1. Metropolitan/nonMetropolitan PSU status.
2. NAEP region (see Section 2.2 for a definition of NAEP region).
3. Public/Catholic/nonCatholic private status.
4. High minority status: whether or not the school has greater than 15 percent minority students.
5. College-bound status: whether the school has greater than 50 percent students who will go on to college.

Nonresponse cells were defined based on cross-classifications of these school and PSU characteristics. The cells were defined as having responding sample sizes greater than 15, with as much difference in response rates between cells as is possible. Cells with small differences in nonresponse rates were collapsed, whether or not they satisfied the 15 sample size minimum.

The nonresponse cells were chosen using a CHAID analysis to define cells with a maximum degree of heterogeneity in response rate across cells. Heterogeneity across cells is equivalent to homogeneity within cells.

CHAID is the name given to one version of the Automatic Interaction Detector (AID) that has been developed for categorical variables. Kass (1980) presents the theory underlying the CHAID technique. The CHAID methodology creates a cell structure based on splitting the data set progressively in a tree structure. The iterative splitting along each newly created branch is done by choosing the "best" variable which has not yet been used on that branch, using modified χ^2 tests. The χ^2 tests are modified using Bonferroni type adjustments to prevent variables from being 'favored' simply because they have more categories. Based on this technique, a 25 percent significance level was required for the χ^2 tests, and a minimum cell size of 15 was assigned.

6.5.3 The School Nonresponse Cells: Results of the CHAID Analysis

The CHAID analysis was carried out using unweighted response rates. Of the 379 schools in the HSTS sample, 340 participated in the HSTS survey, achieving a response rate of 89.7 percent. The analysis was carried out using the five characteristics indicated in Section 6.5.2, with response status as the binary dependent variable. Polychotomous variables such as NAEP Census region were not combined into coarser categories, as is an option with CHAID. The best primary variable in terms of heterogeneity of response was found to be public/Catholic/non-Catholic private status. The counts of schools and response rates are given in Table 6-6.

Table 6-6. Response rates for public, Catholic, and non-Catholic private schools

School Type	Total HSTS sample schools	Response rate by type of school
Public	332	91.9
Catholic	14	92.9
Non-Catholic private	33	66.7
Total	379	89.7

The Catholic school sample consisted of 14 schools, one less than our designated minimum of 15. Nonetheless, the category of Catholic schools was chosen as one of the final nonresponse cells given its importance and the closeness of its sample size to the lower bound. The non-Catholic private schools were further broken out into two cells based on college-bound status.

The public schools were broken out into four branches based on NAEP region. Two of these NAEP region groupings were divided into two cells. Northeast region schools were broken out by minority status, and Southeast region schools were broken out by Metropolitan PSU status.

There were a total of nine nonresponse cells defined across the three types of schools. Table 6-7 presents these cells, the total count of HSTS schools in each cell, and the response rates within the cells.

Table 6-7. Response rates for the school nonresponse cells

School nonresponse cell	Number of HSTS sample schools	Response rate
Private		
Catholic	14	92.9
Non-Catholic private low college-bound	18	50.0
Non-Catholic private high college-bound	15	86.7
Public		
Northeast region, low minority status	27	81.5
Northeast region, high minority status	29	96.5
Southeast region, nonmetropolitan	32	93.8
Southeast region, metropolitan	45	100.0
Central region	83	88.0
West region	116	92.2

6.5.4 HSTS School Nonresponse Adjustments

The HSTS school nonresponse adjustments are computed using the school nonresponse cells selected from the CHAID analysis. The nonresponse adjustments are the reciprocals of weighted response rates computed for each cell. The weights used in these weighted response rates are the numbers of 12th-graders in each school, divided by the probability of selection of the school.

The school base weight, which is the reciprocal of the overall probability of selecting the j^{th} school in the i^{th} PSU, is:

$$W_{ij} = W_i W_{j|i}$$ (Equation 6.5.4)

The school nonresponse adjustment factor for the HSTS weights is designated SCNRAF. It is computed for the α^{th} school nonresponse cell as follows:

$$SCNRAF_{\alpha} = \frac{\sum_{ij \in S(\alpha)} w_{ij} G_{ij}}{\sum_{ij \in SR(\alpha)} w_{ij} G_{ij}} \qquad \text{(Equation 6.5.5)}$$

The subscript ij indicates school j in PSU i.

$SCNRAF_{\alpha}$ denotes the school nonresponse adjustment factor for all schools in the α^{th} school nonresponse adjustment class.

$S(\alpha)$ is the set of all eligible sample schools in the HSTS sample in the α^{th} school nonresponse adjustment class. If a substitute school is used, it replaces the original school in this set.

$SR(\alpha)$ is the set of all schools in the α^{th} school nonresponse adjustment class which have cooperated with the HSTS survey.

G_{ij} is the 12th grade enrollment for the j^{th} school in the i^{th} PSU.

Table 6-8 presents the final school nonresponse factors for each of the nine school nonresponse cells, as computed using Equation 6.5.5.

Table 6-8. Final HSTS school nonresponse factors by nonresponse cell

School Nonresponse Adjustment Cell	Number of HSTS sample schools	Total weighted student count	Schools cooperating in HSTS	Total weighted student count	School nonresponse adjustment factors (SCNRAF)
Catholic	14	114.6	13	112.4	1.020
Non-Catholic private					
Low college bound	18	50.9	9	21.7	2.342
High college bound	15	57.3	13	46.2	1.240
Public Northeast					
Low minority	27	394.6	22	314.2	1.256
High minority	29	219.7	28	211.6	1.038
Public Southeast					
Metropolitan	32	288.6	30	269.8	1.070
Nonmetropolitan	45	436.2	45	436.2	1.000
Public other					
Central	83	698.7	73	611.1	1.143
West	116	959.7	107	879.1	1.092
Total	379	3,220.3	340	2,902.3	1.110

The columns of Table 6-8 are as follows:

1. HSTS sample schools: the counts of schools in $S(\alpha)$.
2. Total weighted student count: the summation of $W_{ij}G_{ij}$ over $S(\alpha)$, given in thousands.
3. Schools cooperating in HSTS Study: the count of schools in $SR(\alpha)$.
4. Total weighted student count: the summation of $W_{ij}G_{ij}$ over $SR(\alpha)$, given in thousands.
5. School nonresponse adjustment SCNRAF, as computed using Equation 6.5.5. These nonresponse factors, as well as the nonresponse factors in the Table 6.5, are computed from the unrounded weight totals. They are not necessarily equal to the ratio of the rounded weight totals given in the tables.

6.5.5 School Nonresponse Adjustment for the NAEP-HSTS Linked Weights

The difference in the school nonresponse adjustment for linked weights with the corresponding adjustment for the HSTS weights is due to the smaller set of responding schools in the former case. We designate as responding schools only those schools which were assigned the particular assessment session type in question, that cooperated with the NAEP assessment, and that sent us transcripts for the HSTS Study.

The school nonresponse cells selected in the CHAID analysis discussed in Section 6.5.2 were also used for the linked weights. The differences in response rates and responding sample sizes should be negligible, so nonresponse cells which are found to have the desired properties for the HSTS weights should also have the same properties with linked weights.

The school weight, which is the reciprocal of the overall probability of selecting the j^{th} school in the i^{th} PSU, is:

$$w_{ij} = w_i w_{j|i}$$

The school nonresponse adjustment factor for the excluded student linked weights will be designated $SCNRFL_\alpha$. It is computed for the α^{th} school nonresponse cell as follows:

$$SCNRFL_\alpha = \frac{\sum_{ij \in SL(\alpha)} w_{ij} G_{ij}}{\sum_{ij \in SRL(\alpha)} w_{ij} G_{ij}} \qquad \text{(Equation 6.5.6)}$$

where

$SCNRFL_\alpha$ denotes the school nonresponse adjustment factor for all linked schools in the α^{th} school nonresponse adjustment class.

$SL(\alpha)$ is the set of all eligible sample schools in the HSTS sample in the α^{th} school nonresponse adjustment class. Substitute schools are not included in this set. This set is the same as $S(\alpha)$ from Section 6.5.4.

$SRL(\alpha)$ is the set of all schools in the α^{th} school nonresponse adjustment class which have cooperated with the HSTS survey, and have also responded in the NAEP assessment.[19] This set should be smaller than the corresponding $SR(\alpha)$ from Section 6.5.4.

G_{ij} is the 12th grade enrollment for the j^{th} school in the i^{th} PSU.

Table 6-9 presents the school nonresponse adjustment factors computed for each of the nine school nonresponse cells as computed by Equation 6.5.6. The weighted totals are given in thousands.

Table 6-9. HSTS-NAEP school nonresponse factors by nonresponse cell

School Nonresponse Adjustment Cell	Number of HSTS sample schools	Total weighted student count	Schools participating in NAEP and HSTS studies	Total weighted student count	School nonresponse adjustment factors for excluded students (SCNRFL)
Catholic	14	114.6	13	112.4	1.020
Non-Catholic Private					
Low college-bound	18	50.9	8	16.0	3.173
High college-bound	15	57.3	13	46.2	1.240
Public Northeast					
Low minority	27	394.6	18	255.1	1.547
High minority	29	219.7	25	191.1	1.150
Public Southeast					
Metropolitan	32	288.6	25	218.8	1.319
Nonmetropolitan	45	436.2	41	389.4	1.120
Public Other					
Central	83	698.7	55	447.1	1.563
West	116	959.7	82	641.5	1.496
Total	379	3,220.3	280	2,317.6	1.389

[19] This set excludes three schools which cooperated with both surveys, but could not provide information linking transcripts to the assessed students. In effect, a separate transcript sample was drawn as if the school was a NAEP nonrespondent.

The columns of Table 6-9 are as follows:

1. HSTS sample schools: the count of schools in *SL(a)*.
2. Total weighted student count: summation of $W_{ij}G_{ij}$ over *SL(a)*, given in thousands.
3. Schools cooperating in NAEP and HSTS Studies: the count of schools in *SRL(a)*.
4. Total weighted student count: the summation of $W_{ij}G_{ij}$ over *SRL(a)*.
5. SCRNFL: the school nonresponse adjustment for the cell, as computed in Equation 6.5.6.

For each nonresponse cell, the *SCNRFL* value is greater than or equal to the corresponding *SCNRAF* value.

The school nonresponse adjustment factor for the linked weights for each assessment a will be slightly different from $SCNRFL_{\alpha}$ and is designated $SCNRFL_{a\alpha}$. It was computed for the α^{th} school nonresponse cell as follows:

$$SCNRFL_{a\alpha} = \frac{\sum_{ij \in SL_a(\alpha)} w_{ij} w_{a|ij} G_{ij}}{\sum_{ij \in SRL_a(\alpha)} w_{ij} w_{a|ij} G_{ij}} \qquad \text{(Equation 6.5.7)}$$

where

$SCNRFL_{a\alpha}$ denotes the school nonresponse adjustment factor for the a^{th} assessment for all schools in the α^{th} school nonresponse adjustment class.

$SL_a(\alpha)$ is the set of all eligible sample schools in the HSTS sample who were also assigned the a^{th} assessment, in the α^{th} school nonresponse adjustment class. Substitute schools are not included in this set.

$SRL_a(\alpha)$ is the set of all schools in the α^{th} school nonresponse adjustment class that responded in the NAEP assessment, were assigned to the a^{th} assessment, and participated in the HSTS survey.

$w_{a(j)}$ is the inverse of the conditional probability that at least one session of the assessment in question has been assigned to school ij (see Section 6.4.3). This quantity is equal to 1 for most schools, but will be equal to 2 for smaller schools which had only a reading or a history/geography assessment.

Tables 6-10 and 6-11 present these nonresponse adjustment factors (computed from Equation 6.5.7) for the reading and history/geography assessments.

The columns in the two tables are as follows:

1. HSTS-NAEP assessment sample schools: the count of schools in $SL(a)$.
2. Total weighted students count: the summation of $W_{ij}G_{ij}$ over $SL(a)$, given in thousands.
3. Assessment schools cooperating in both NAEP and HSTS Studies: the count of schools in $SRL_{\alpha(\alpha)}$
4. Total weighted student count: the summation of $W_{ij}G_{ij}$ over $SRL_{\alpha(\alpha)}$, given in thousands.
5. Assessment $SRL_{\alpha(\alpha)}$: the school nonresponse adjustment for the cells as computed using Equation 6.5.7.

Table 6-10. HSTS-NAEP reading assessment school nonresponse factors

School Nonresponse Adjustment Cell	Number of HSTS NAEP reading assessment sample schools	Total weighted student count	Reading assessment schools participating in both NAEP and HSTS studies	Total weighted student count	Reading assessment SCNRFL
Catholic	14	116.8	13	112.4	1.039
Non-Catholic Private					
Low college-bound	10	49.2	4	14.2	3.463
High college-bound	10	57.4	9	48.1	1.193
Public Northeast					
Low minority	27	394.6	18	255.1	1.547
High minority	29	219.7	25	191.1	1.150
Public Southeast					
Metropolitan	32	288.6	25	218.8	1.319
Nonmetropolitan	45	436.2	41	389.4	1.120
Public other					
Central	81	706.1	53	451.2	1.565
West	112	964.9	78	646.6	1.492
Total	360	3,233.5	266	2,326.9	1.390

Table 6-11. HSTS-NAEP history/geography assessment school nonresponse factors

School Nonresponse Adjustment Cell	Number of HSTS NAEP history/ geography assessment sample schools	Total weighted student count	History/Geography assessment schools cooperating in both NAEP and HSTS studies	Total weighted student count	History geography assessment SCNRFL
Catholic	13	112.4	13	112.4	1.000
Non-Catholic Private					
Low college-bound	13	52.5	5	17.9	2.942
High college-bound	11	57.1	9	44.3	1.291
Public Northeast					
Low minority	27	394.6	18	255.1	1.547
High minority	29	219.7	25	191.1	1.150
Public Southeast					
Metropolitan	32	288.6	25	218.8	1.319
Nonmetropolitan	45	436.2	41	389.4	1.120
Public Other					
Central	77	691.2	51	443.1	1.560
West	110	954.6	76	636.4	1.500
Total	357	3,206.9	263	2,308.5	1.389

6.6 Student Nonresponse Adjustments

The final weight for each student is the base weight multiplied by a number of special factors These factors in their usual order of implementation are as follows:

1. An adjustment for nonresponse at the school level.
2. An adjustment for nonresponse of the student to a NAEP assessment.
3. An adjustment for missing transcripts.
4. An adjustment for 'large' weights (trimming).
5. An adjustment to known CPS student population totals (poststratification).

We note that this is the "usual" order of implementation for weighting in surveys of this kind (such as 1994 NAEP), but the actual implementation in 1994 HSTS put the adjustment of missing transcripts at the end, for reasons discussed below. The adjustment for nonresponse at the school level was discussed in Section 6.5. We also need to adjust the weights for nonresponse at the student level. These adjustments are discussed in Section 6.6.1. In general practice, adjustment for poststratification is the last step, since we generally desire the final weights to aggregate exactly to the poststratification control totals. (Any adjustment following the poststratification step will cause the final weights not to satisfy this property.) Any nonresponse adjustments are computed first, followed by a trimming adjustment for large weights, followed by the final poststratification step to generate weights that aggregate exactly to known control totals.

In the 1994 HSTS, however (as in the 1990 HSTS Study) we decided to make an adjustment for missing transcripts follow the poststratification step (see Section 6.7). The other nonresponse adjustments, including the adjustments for students who did not complete an assessment, precede the trimming and poststratification step, as is general practice.

There were several reasons for making the missing transcripts adjustment the final step. First, the nonresponding students were, for the most part, nonrespondents only in the sense that a transcript was not collected for them. For the large majority of such students, data were collected on their race/ethnicity and age -- characteristics needed for poststratification. This information made it possible to include these students in the derivation of poststratification factors. Second, the missing transcript nonresponse adjustments were applied only to graduates, whereas the poststratification factors were derived using both a population and a sample of 12th-graders that included some nongraduating students. The nonresponse adjustments for students not completing assessments, on the other hand, do include nongraduating grade 12 students. Finally, the adjustment for missing transcripts is fairly small, so the deviation of the aggregated final weights from the control totals is negligible.

The details of the missing transcript adjustments are discussed in Section 6.7.2. The trimming adjustments are discussed in Sections 6.6.2 and 6.6.3. The poststratification adjustments are discussed in Sections 6.6.4 through 6.6.6.

6.6.1 Student Nonresponse Adjustments for Assessed Students

Within each school, samples were drawn of the 12th-grade students who were then randomly assigned to assessments. Any student found to be ineligible at this point was excluded from an assessment. Many of the students assigned to assessments did not actually take an assessment exam, either because of a refusal to participate or because of an absence on the day of the assessment. This section discusses adjustments made in the linked weights for this student level assessment nonresponse.

As we discussed in Section 6.5, nonresponse is a concern in any study because of the possibility that the study results will be invalidated by nonresponse bias. Bias could be incurred from a lack of participation from a subset of students, because this group will be "self-selected." The 1994 NAEP assessment made adjustments to lower this bias using nonresponse adjustments within a selected group of nonresponse cells. The 1994 HSTS Study used the same nonresponse cells and the same methodology for determining nonresponse adjustments. However, the actual nonresponse adjustments for the two studies differ because the set of schools selected for the HSTS study was only a subset of the original set of schools participating in the NAEP assessment.

The nonresponse cells for HSTS are the same as were used for NAEP. The NAEP nonresponse cells are based on the NAEP PSU sampling strata and the age and race of the student. The PSU sampling strata are grouped into stratum groupings for these cells (this grouping is slightly different for reading assessment students and history/geography assessment students). A dichotomous age status was used for generating nonresponse cells, indicating whether the student was born on or before September 30, 1975 or the student was born later. A trichotomous race status was used for generating nonresponse cells, with the first category white or Asian; the second category black, Hispanic, or other; and the third category missing race status.

In the 1994 NAEP study, nonresponse adjustments were made for the excluded students without completed excluded questionnaires. These adjustments were not made for excluded students in the 1994 HSTS weights, however, because even without the questionnaire information, we obtained most of the information for these students that would be of interest to analysts of the HSTS data.

We will indicate as $ST_a(\gamma)$ the set of all students assigned to the a^{th} assessment (reading, history, or geography) in the γ^{th} student nonresponse cell, and define $STR_a(\gamma)$ as the corresponding set of students who actually completed the a^{th} assessment. There were 51 student

nonresponse cells in all, defined slightly differently for reading session students and history/geography session students.

If we define $STNNRF_{a\gamma}$ as the student nonresponse adjustment factor for the a^{th} assessment and the γ^{th} student nonresponse cell, then Equation 6.6.1 below indicates how these quantities are computed.

$$STNNRF_{a\gamma} = \frac{\sum_{ijk \in ST_a(\gamma)} w^a_{ijk} SCNRFL_{aa} I[ij \in SRL_a(\alpha)]}{\sum_{ijk \in STR_a(\gamma)} w^a_{ijk} SCNRFL_{aa} I[ij \in SRL_a(\alpha)]} \qquad \text{(Equation 6.6.1)}$$

The quantity w^a_{ijk} is the student base weight for assessment a assigned to the k-th student in the j-th school in the i-th PSU, as discussed in Section 6.4.3. The quantity $SCNRFL_{aa}$ is the assessment a school nonresponse adjustment computed for school ij, discussed in Section 6.5.5. The indicator function $I[ij \in SRL_a(\alpha)]$ is equal to 1 if school ij is in school nonresponse cell α, and equal to 0 otherwise.

We also need a special nonresponse adjustment when we are computing poststratification adjustments for the excluded student weights. These poststratification adjustments pool all of the assessed students, regardless of assessment, and all of the excluded students into one group. The nonresponse cells which will be used are the reading session nonresponse cells for the reading session students, and the history/geography session nonresponse cells for the history/geography session students, a total of 102 cells. The special nonresponse adjustment factor for students within these cells is given in Equation 6.6.2.

$$STNNRF_{\gamma} = \frac{\sum_{ijk \in ST(\gamma)} w_{ijk} SCNRFL_{a} I[ij \in SRL(\alpha)]}{\sum_{ijk \in STR(\gamma)} w_{ijk} SCNRFL_{a} I[ij \in SRL(\alpha)]} \qquad \text{(Equation 6.6.2)}$$

The set $ST(\gamma)$ represents all assessed students (of any assessment) in the γ^{th} student nonresponse cell. The set $STR(\gamma)$ corresponds to the assessed students who were successfully assessed. The quantity w_{ijk} is the base weight of the student, including only the school base weight and the inverse of the probability of selection of the student into the NAEP sample. In other words, the base weight does not include probabilities of selection into separate assessments (see Section 6.4). The school nonresponse adjustment ($SCNRFL_a$) used here also does not distinguish between

assessments: all schools with any assessment are included in the computation of this factor (see Section 6.5.5). The indicator function $I[ij \in SRL(\alpha)]$ is equal to 1 if school ij is in school nonresponse cell α, and equal to 0 otherwise.

Table 6.12 presents percentiles for the student nonresponse adjustments $STNNRF_{\alpha\gamma}$ for the three assessments, and the special nonresponse adjustment $STNNRF_{\gamma}$.[20] There are 51 unique values for each of the assessment adjustments and 102 unique values for the excluded student nonresponse adjustment. The minimum and maximum values of these values is given for each adjustment in the table. In addition, the weighted p-th percentile is given for the 10th, 25th, 50th, 75th, and 90th percentiles. The weighted 10th percentile, for example, is that value of the nonresponse adjustment for which a subset of responding assessed students with a smaller or equal adjustment, correspond to 10 percent of the weights. In other words, if the 10th percentile for the reading assessment nonresponse adjustment is 1.058, then 10 percent of the weight corresponds to responding reading assessment students having nonresponse adjustments that are less than or equal to 1.058. The mean value is the average of the student nonresponse adjustment over all students in that particular category. Note that the excluded student nonresponse adjustment percentiles are over all students who were assessed or excluded, regardless of assessment.

Table 6-12. Student nonresponse adjustments for reading, history, and geography assessments and for excluded students by percentile

Percentile	Type of Assessment			
	Reading	History	Geography	Excluded
Minimum	1.019	1.043	1.000	1.019
10th	1.058	1.089	1.065	1.065
25th	1.129	1.116	1.119	1.130
50th (median)	1.194	1.224	1.186	1.204
75th	1.287	1.285	1.287	1.281
90th	1.326	1.352	1.378	1.364
Maximum	1.436	1.609	1.485	1.454
Mean	1.205	1.220	1.213	1.211

[20] Although $STNNRF_{\gamma}$ is used in calculating the final weights for excluded students, data from all assessed and excluded students are needed to calculate $STNNRF_{\gamma}$.

6.6.2 Trimming the Nonresponse Adjusted Student Weights

The students in some schools were assigned extremely large weights because the school was predicted (on the basis of the QED data) to have a small number of eligible students, yet in fact had a large number. Other excessively large weights may result from differential response rates. To reduce the effect of large contributions to variance from a small number of schools, the weights of such schools were reduced or "trimmed." The trimming procedure may introduce a small bias but is designed to reduce the mean square error of sample estimates.

The trimming algorithm is identical to the one that Westat has used for all recent NAEP survey weights (including the 1994 NAEP weights). The algorithm has the effect of trimming the overall weight of any school that contributes more than a specified proportion θ to the estimated variance of the estimated number of students eligible for the HSTS Survey.

The trimming algorithm described in this section defines the trimming adjustments for the HSTS weights. Let M be the number of responding HSTS schools in the sample. Define *SCHR(ij)* as the set of students who were included in the HSTS survey in school *ij*. Define

$$x_{ij} = \sum_{ijk \in SCHR(ij)} w_{ijk}\, SCNRAF_{\alpha}\, I[ij \in SR(\alpha)] \qquad \text{(Equation 6.6.3)}$$

The two factors incorporating the school nonresponse adjustment are discussed in Section 6.5.4. The quantity x_{ij} is the sum of the school nonresponse adjusted student base weights in the school. Define *SR* as the overall set of schools cooperating with the HSTS survey, and define

$$\bar{x} = \frac{1}{M} \sum_{ij \in SR} x_{ij} \qquad \text{(Equation 6.6.4)}$$

$\bar{x}$ is the mean value of the x_{ij}'s over all participating HSTS schools. The following sum of squares will be used in our trimming procedure:

$$V = \sum_{ij \in SR} (x_{ij} - \bar{x})^2 \qquad \text{(Equation 6.6.5)}$$

If any school contributes too large a share to this sum of squares, then the school and student weights will be contributing significantly to the sampling variance of most estimators. We will

impose as a constraint the following requirement: for each school $lm \in SR$ such that $x_{lm} > \bar{x}$ we require that

$$(x_{lm} - \bar{x})^2 \leq \theta \sum_{ij \in SR} (x_{ij} - \bar{x})^2 \qquad \text{(Equation 6.6.6)}$$

We selected the value of θ based on empirical experience in surveys such as NAEP. This value is *10/M*.

In order to impose this requirement, an iterative trimming procedure is carried out on the student weights. The first step is to compute

$$\theta_{ij}(1) = \frac{(x_{ij}(1) - \bar{x}(1))^2}{V(1)} \qquad ij \in SR \qquad \text{(Equation 6.6.7)}$$

The argument "*1*" indicates that these are the values of these quantities preceding the first iteration of the trimming procedure. If no value of $\theta_{ij}(1)$ exceeds *10/M*, then trimming is unnecessary. If at least one value of $\theta_{ij}(1)$ exceeds *10/M* (with $x_{ij}(1)$ also exceeding $\bar{x}(1)$), then choose $lm \in SR$ such that $\theta_{lm}(1)$ exceeds $\theta_{ij}(1)$ for all *ij* not equal to *lm*, and such that $x_{lm}(1)$ also exceeds $\bar{x}(1)$. For this school we will compute an adjusted school base weight $w_{lm}(2)$ which is equal to

$$w_{lm}(2) = w_{lm}(1)\left[\frac{\bar{x}(1)}{x_{lm}(1)} + \sqrt{\frac{10/M}{\theta_{lm}(1)}}\left|1 - \frac{\bar{x}(1)}{x_{lm}(1)}\right|\right] \qquad \text{(Equation 6.6.8)}$$

$w_{lm}(1)$ is equal to the original base weight w_{lm}. After this computation, carry out the following steps:

1. Recompute x_{lm} as:

$$x_{lm}(2) = \sum_{lmk \in SCHR(lm)} w_{lm}(2) w_{k|lm} SCNRAF_{\alpha} I[lm \in SR(\alpha)] \qquad \text{(Equation 6.6.9)}$$

2. Reassign $x_{ij}(2) = x_{ij}(1)$ for all $ij \in SR$ not equal to *lm*.

3. Recompute $\bar{x}(2)$ and *V(2)*.

At this point, the first iteration is completed. Suppose *t-1* iterations have been completed (*t*=2,....). Then the *t-th* iteration will have the following steps:

1. Recompute the θ_{ij}:

$$\theta_{ij}(t) = \frac{(x_{ij}(t) - \bar{x}(t))^2}{V(t)} \qquad ij \in SR \qquad \text{(Equation 6.6.10)}$$

2. If no value of $\theta_{ij}(t)$ exceeds *10/M* then further trimming will be unnecessary (all schools now satisfy the constraint). The trimming algorithm is complete.

3. If at least one value of $\theta_{ij}(t)$ exceeds *10/M* (with $x_{ij}(t)$ also exceeding $\bar{x}(t)$) then choose $lm \in S$ such that $\theta_{lm}(t)$ exceeds $\theta_{ij}(t)$ for all *ij* not equal to *lm* and such that $x_{lm}(t)$ also exceeds $\bar{x}(t)$. For this school we will compute an adjusted school base weight $w_{lm}(t+1)$ which will be equal to

$$w_{lm}(t+1) = w_{lm}(t)\left[\frac{\bar{x}(t)}{x_{lm}(t)} + \sqrt{\frac{10/M}{\theta_{lm}(t)}}\left|1 - \frac{\bar{x}(t)}{x_{lm}(t)}\right|\right] \qquad \text{(Equation 6.6.11)}$$

In general, $w_{lm}(t)$ will be equal to the original school base weight w_{lm}, unless the school's weight was trimmed in an earlier iteration. The final steps of the iteration are as follows:

1. Recompute x_{lm} as:

$$x_{lm}(t+1) = \sum_{lmk \in SCHR(lm)} w_{lm}(t+1) w_{k|lm}\, SCNRAF_{\alpha}\, 1[lm \in SR(\alpha)] \qquad \text{(Equation 6.6.12)}$$

2. Reassign $x_{ij}(t+1) = x_{ij}(t)$ for all $ij \in SR$ not equal to *lm*.

3. Recompute $\bar{x}(t+1)$ and *V(t+1)*.

This ends the t-th iteration. These iterations are continued until there is no further trimming to be done -- that is, until all adjusted weights satisfy the criterion. Suppose *T* is the final iteration and $x_{ij}(T)$ the final school weight for each school *ij*. We compute a trimming factor *TRIM(ij)* for each school equal to:

$$TRIM(ij) = \frac{x_{ij}(T)}{x_{ij}(1)} \qquad \text{(Equation 6.6.13)}$$

Trimming was necessary for only three of the schools in the HSTS sample. The final trimming factors for these schools were 0.576, 0.770, and 0.891.

6.6.3 Trimming the Linked Base Weights

Trimming was also carried out on the school and student nonresponse adjusted link weights. The algorithm used was identical to that discussed in Section 6.6.2. Trimming factors were computed for each school *ij* for the school and student nonresponse adjusted linked base weights w^a_{ijk} (for each assessment *a*), and for the school and student nonresponse adjusted linked base weights w^e_{ijk} (for excluded students).

For the assessment weights the set of schools that are included in the trimming computations are designated SRL_a. These include for each assessment all schools that responded in the NAEP assessment, were assigned to the *a-th* assessment, and participated in the HSTS survey. For the excluded student weights, the set of schools that are included in trimming computations is the set *SRL*. This set includes all schools that participate in the NAEP assessment and the HSTS survey, regardless of assessment assignments.

For the HSTS weights, the inputs to the trimming algorithm were the summations of nonresponse adjusted base weights over all students for each school ij: the x_{ij}. For the assessment *a* base weights the corresponding inputs are as follows:

$$x^a_{ij} = \sum_{ijk \in SCHR(ij),\ ijk\ assessed} w^a_{ijk}\ SCNRFL_{a\alpha}\, I[ij \in SRL_a(\alpha)]\ STNNRF_{a\gamma}\, I[ijk \in STR_a(\gamma)] + \sum_{ijk \in SCHR(ij),\ ijk\ excluded} w^e_{ijk}\ SCNRFL_a\, I[ij \in SRL(\alpha)] \qquad \text{(Equation 6.6.14)}$$

For each term in the two right hand summations the second and third factors incorporate the school nonresponse adjustment (see Section 6.5), and the fourth and fifth factors in the first summation incorporate the student assessment nonresponse adjustment factor (see Section 6.6.1). These x^a_{ij} quantities are computed for all schools in SRL_a. The trimming factors for these schools at the end of the algorithm are designated as $TRIM_a(ij)$.

Trimming factors need also to be computed for the special weights to generate excluded student weights. The excluded students receive separate poststratification adjustments in their base weights. These adjustments, however, include all students, including all of the assessed students, since the control totals include all students (see Section 6.7 for details). The trimming algorithm is therefore run trimming these aggregated weights for each school. The input factors for these special adjustments are as follows:

$$x_{ij}^{S} = \sum_{ijk \in SCHR(ij), ijk\ assessed} w_{ijk}\ SCNRFL_{\alpha} I[ij \in SRL(\alpha)]\ STNNRF_{\gamma}\ I[ijk \in STR(\gamma)] + \sum_{ijk \in SCHR(ij), ijk\ excluded} w_{ijk}\ SCNRFL_{\alpha} I[ij \in SRL(\alpha)] \qquad \text{(Equation 6.6.15)}$$

The trimming factors generated from the algorithm using these inputs are designated $TRIM_s(ij)$.

The same three schools that needed trimming for the HSTS weights also needed trimming on at least one of the linked weights. Table 6-13 presents these trimming factors for the HSTS weights ($TRIM(ij)$), for each of the three assessment weights ($TRIM_a(ij)$), and the special weight for excluded students ($TRIM_s(ij)$). A trimming factor of 1 indicates that the weight did not require trimming.

Table 6-13. Trimming factors for schools requiring trimming

School (NAEP identifier)	HSTS trimming factor	Reading trimming factor	History trimming factor	Geography trimming factor	Special trimming factor
104330	0.576	0.689	0.759	0.684	0.696
512333	0.770	0.854	0.816	0.801	0.819
514330	0.891	1.000	1.000	0.843	1.000

6.6.4 Poststratified Student Weights

In most sample surveys, the respondent weights are random variables that are subject to sampling variability. Even if there were 100 percent response, the respondent weights would at best provide unbiased estimates of the various subgroup proportions. However, since unbiasedness refers

to average performance over a conceptually infinite number of replications of the sampling, it is unlikely that any given estimate, based on the achieved sample, will exactly equal the population value. Furthermore, the respondent weights have been adjusted for nonresponse and a few extreme weights have been reduced in size.

To reduce the mean square error of estimates using the sampling weights, these weights will be further adjusted so that estimated population totals for a specified subgroup population, based on the sum of student weights for a specified type, will be the same as presumably better estimates based on composites of estimates from the Current Population Survey. This adjustment, called poststratification, is intended especially to reduce the mean squared error of estimates relating to student populations that span several subgroups of the population. The poststratification classes are defined in terms of race/ethnicity and NAEP region.

For the HSTS weights, the post-stratification adjustment factor ($STPSAF_g$) for the g^{th} post-stratification adjustment cell will be:

$$STPSAF_g = \frac{C_g}{\sum_{ijk \in E(g)} w_{ijk}\, SCNRAF_\alpha I[ij \in SR(\alpha)]\, TRIM(ij)} \qquad \text{(Equation 6.6.16)}$$

The quantity C_g is the 12th grade enrollment control total of students whose 18th birthday was on or after January 1, 1994 for the g^{th} poststratification class. E(g) is the collection of all students in the g^{th} poststratification class who were enrolled in 12th grade (including those who did not graduate in 1994) and whose 18th birthday was on or after January 1, 1994. The quantity w_{ijk} is the full sample student base weight for the k^{th} student in the j^{th} school in the i^{th} PSU, that was discussed in Section 6.4.1. The final three factors comprise the school nonresponse adjustment for the HSTS weights, discussed in Section 6.5.4., and the trimming factor for the school, discussed in Section 6.6.2.

Table 6-14 presents the poststratification cells with the CPS control totals for each cell. Control totals are given in thousands. For a discussion of the definition of regions as used in NAEP, see Section 2.2.

Table 6-14. Student poststratification cells and control totals

Poststratification cell	Race/Ethnicity	Region	CPS control total (000)
1	Blacks, nonHispanic	All	235.3
2	Hispanics	All	159.2
3	Other race, nonHispanic	All	102.6
4	Whites, nonHispanic	Northeast	347.0
5	Whites, nonHispanic	Southeast	342.8
6	Whites, nonHispanic	Central	494.7
7	Whites, nonHispanic	West	414.5

Table 6-15 presents the aggregated weights within each poststratification cell (the denominator of Equation 6.6.16), the control total C_g, and the poststratification factor $STPSAF_g$ for the poststratification cell.

Table 6-15. HSTS poststratification factors

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification factor
1	166.2	235.3	1.416
2	144.8	159.2	1.099
3	105.1	102.6	0.976
4	287.1	347.0	1.209
5	255.7	342.8	1.341
6	314.4	494.7	1.573
7	266.3	414.5	1.557

In Table 6-15 and the remaining tables in Section 6.6, the poststratification factor as given is the unrounded control total divided by the unrounded aggregated weight. The control totals and aggregated weights given in the tables are the corresponding total rounded to one digit after the decimal point. The poststratification factor as given may not equal the ratio of the two rounded summands as given in all cases.

6.6.5 Poststratification for the Linked Weights: Assessment Weights

The poststratification procedure is similar to the corresponding procedure for the HSTS weights as described in Section 6.6.4, in that the same poststratification categories and control totals are used. In this case, however, separate adjustments are made for each of the three assessments, and for the excluded students.

For the three assessments, each assessment sample must represent the full population. The control totals however are not separable into students eligible for an assessment, and excluded students. Because of this nonseparability, the excluded students from the sample must be included with the assessment group when computing the poststratification adjustment. For each assessment a the poststratification factor corresponding to poststratification class g is as follows:

$$STPSFL_{ag} = \frac{C_g}{\left[\sum_{ijk \in E(g),\ ijk\ assessed} w^a_{ijk}\ SCNRFL_{a\alpha}\, I[ij \in SRL_a(\alpha)]\ STNNRF_{a\gamma}\, I[ijk \in STR_a(\gamma)]\ TRIM_a(ij) + \sum_{ijk \in E(g),\ ijk\ excluded} w^e_{ijk}\ SCNRFL_{\alpha}\, I[ij \in SRL(\alpha)]\ TRIM_a(ij) \right]}$$

(Equation 6.6.17)

The quantity C_g in the numerator of Equation 6.6.17 represents the 12th-grade enrollment control total of students whose 18th birthday was on or after January 1, 1994 for the g^{th} poststratification class. E(g) is the collection of all students in the g^{th} poststratification class who were enrolled in 12th-grade (including those who did not graduate in 1994) and whose 18th birthday was on or after January 1, 1994. The quantities w^a_{ijk} and w^e_{ijk} are the student base weights for assessed and excluded students respectively, discussed earlier in Section 6.4.3.

There are school nonresponse adjustment factors in both the assessed and excluded student summations, discussed in Section 6.5.5, and student nonresponse adjustment factors for the assessed students only, discussed in Section 6.6.1. The final factors in each term of each summation are trimming factors for the weights, discussed in Section 6.6.3.

Tables 6-16, 6-17, and 6-18 present the aggregated weights (the denominator of Equation 6.6.17), the control totals C_g, and the poststratification factors $STPSFL_{ag}$ for each poststratification cell for the reading assessment, the history assessment, and the geography assessment, respectively.

Table 6-16. Poststratification factors for the reading assessment weights

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification factor (000)
1	163.6	235.3	1.438
2	149.4	159.2	1.065
3	115.7	102.6	0.886
4	302.9	347.0	1.146
5	245.3	342.8	1.397
6	318.0	494.7	1.556
7	264.0	414.5	1.570

Table 6-17. Poststratification factors for the history assessment weights

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification factor (000)
1	168.7	235.3	1.395
2	140.3	159.2	1.134
3	103.3	102.6	0.993
4	312.0	347.0	1.112
5	262.4	342.8	1.306
6	315.1	494.7	1.570
7	250.7	414.5	1.653

Table 6-18. Poststratification factors for the geography assessment weights

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification factor (000)
1	181.6	235.3	1.296
2	142.8	159.2	1.115
3	124.1	102.6	0.826
4	303.1	347.0	1.145
5	260.2	342.8	1.318
6	299.3	494.7	1.653
7	258.8	414.5	1.601

6.6.6 Special Poststratification Adjustments for the Final Excluded Student Weights

The poststratification adjustment for the excluded students needs to include all students, since control totals do not exist for excluded students alone. In this case, all students from all of the assessments are included, along with the excluded students, when computing the adjustments. The weights used for these students are not the weights adjusted for selection into an assessment. Rather, they are the original weights reflecting selection into the HSTS sample: the original HSTS base weights adjusted for school nonresponse (using the excluded student linked weight adjustments).

The poststratification adjustment for excluded students is shown as follows:

$$STPSFL_{eg} = \frac{C_g}{\left[\sum_{ijk \in E(g),\ ijk\ assessed} w_{ijk} SCNRFL_{\alpha} I[ij \in SRL(\alpha)]\ STNNRF_{\gamma}\ I[ijk \in STR(\gamma)]\ TRIM_S(ij) + \sum_{ijk \in E(g),\ ijk\ excluded} w_{ijk} SCNRFL_{\alpha} I[ij \in SRL(\alpha)]\ TRIM_S(ij) \right]} \quad \text{(Equation 6.6.18)}$$

The school nonresponse adjustment factors were discussed in Section 6.5.5, student nonresponse adjustment factors in Section 6.6.1, and trimming factors in Section 6.6.3.

Table 6-19 presents the aggregated weights (the denominator of Equation 6.6.18), the control totals C_g, and the poststratification factors $STPSFL_{eg}$ for each poststratification cell.

Table 6-19. Poststratification factors for the excluded student weights

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification factor (000)
1	168.5	235.3	1.400
2	145.1	159.2	1.097
3	113.7	102.6	0.902
4	306.5	347.0	1.132
5	253.0	342.8	1.355
6	312.6	494.7	1.583
7	259.1	414.5	1.600

6.7 Final Adjustments and Final Sampling Weights

For a small percentage of graduated students it was not possible to obtain a transcript. An adjustment is necessary in the weights of graduated students with transcripts to account for this. In order to do this adjustment correctly, it is necessary to have the complete set of graduated students, with or without transcripts. There are a small set of students, however, for whom no transcripts were received and the graduation status was unknown. Among these students, a certain percentage were imputed as graduating, based on overall percentages of graduating students. The remainder were imputed as non-graduating.

The imputation process was a standard hot-deck imputation (see, for example, Little and Rubin (1987), Section 4.5.3). For each student with a usable transcript and unknown graduation status, a "donor" was randomly selected (without replacement) from the set of all students with usable transcripts and known graduation status from the same school, gender, race/ethnicity, and age status. Race/ethnicity was categorized in the same way as for poststratification. The categories were Hispanics, black nonHispanics, white nonHispanics, and other race nonHispanics. Age status was categorized according to birthdate:

1. "Young" students, whose birthdate followed January 1, 1977.
2. "Age Eligible" students, whose birthdate was between January 1 and December 31, 1976.
3. "Old" students, whose birthdate preceded January 1, 1976.

Each student with known graduation status in a cell in a particular school could be used a maximum of three times as a donor for a student in the same cell in the same school with unknown graduation status. If insufficient donors were available within this school within the cell, then donors were randomly selected from students within the cell from other schools with similar characteristics as the school in question. The cells used to define these "similar" schools are based on the following school characteristics:

1. NAEP region (defined in Section 2.2)
2. Public/Catholic/nonCatholic private status
3. College-bound status of the school (whether or not 50 percent of the graduates go on to college).

For example, if a Catholic school in the Northeast NAEP region with more than 50 percent of its students going on to college did not have enough donors in a particular student cell, then donors were randomly drawn from other schools in this class.

Table 6-20 presents counts of the number of students with known and unknown graduation status, the counts of those with known status who graduated or did not graduate, and the counts of those with unknown status who were imputed as graduating or not graduating.

Table 6-20. Counts and percents of graduating seniors known and imputed

	Known graduation status		Imputed graduation status	
Status	Number of students	Percent of students	Number of students	Percent of students
Not graduating	2,717	9.6	53	10.3
Graduating	25,581	90.4	464	89.7
All seniors	28,298	100.0	517	100.0

Note that the percent of students that was imputed as not graduating (10.3 percent) was higher than the corresponding percent of students confirmed as not graduating. This occurred because the students with unknown graduation status tended to fall into groups with higher percentages of nongraduating students.[21]

6.7.1 CHAID Analysis to Choose Missing Transcript Nonresponse Cells

As with school nonresponse, our approach to nonresponse adjustments for missing transcripts was to choose nonresponse cells for students, and assign nonresponse weighting adjustments that are uniform within each cell. These cells should be homogeneous in terms of response propensity within cells, while being heterogeneous in response propensity across cells. The sample size should not be too small in any one cell, so a minimum responding sample size of 30 will be required for each nonresponse cell.

[21] The percentage of nongraduates among students of unknown graduation status may be even higher than was imputed. In general, graduation status is missing from our records because schools could not provide it. Since providing transcripts of graduation is a major function of American high schools, there is a strong presumption that if a high school does not know a senior's graduation status, that student did not graduate.

The nonresponse cells were chosen after an analysis using CHAID (see Section 6.5.2 for a discussion of CHAID). The predictive variables used included NAEP region, public/Catholic/nonCatholic private status of school, race/ethnicity, and gender. Any graduates missing any of these values were assigned imputed values using a hot-deck procedure.

The CHAID analysis chose 11 cells as nonresponse cells. These cells were homogeneous in response rate within cell, and heterogeneous in response rate between cells. Table 6-21 presents these cells, with counts of students and nonresponse rates.

Table 6-21. Nonresponse adjustment cells for missing transcript adjustments

Cell number	Nonresponse cell	Number of students	Nonresponse rate (in percent)
1	NonHispanic whites, Northeast region	3,589	2.7
2	NonHispanic whites, South and Central regions	8,862	1.3
3	NonHispanic whites, West region	4,211	2.5
4	NonHispanic blacks, older students	1,712	4.7
5	NonHispanic blacks, other	2,645	2.2
6	Hispanics, Northeast region	447	13.9
7	Hispanics, South region	331	1.5
8	Hispanics, Central region	270	17.4
9	Hispanics, West region	2,122	2.3
10	NonHispanic other races	1,750	3.0
11	Missing gender status	76	46.1

The final cell consists of 76 students for whom gender was not recorded. The high rate of missing transcripts among Hispanic students in Regions 1 and 3 is concentrated in seven problem schools. Three of these schools are in Region 1 and four in Region 3. These 7 schools had extremely high missing transcript rates for ALL students, with higher nonresponse for Hispanics than for non-Hispanics (see Table 6-22).

Table 6-22. Comparison of rates of missing transcripts in the worst seven schools in Regions 1 and 3 with the remaining schools in those regions

School set	Hispanic status	Students with missing transcripts	All students	Percent with missing transcripts
Region 1				
Worst three schools	Hispanic	55	80	68.8
	Non-Hispanic	63	205	30.7
All other schools	Hispanic	8	368	2.2
	Non-Hispanic	71	4,555	1.6
Region 3				
Worst three schools	Hispanic	38	120	31.7
	Non-Hispanic	67	362	18.5
All other schools	Hispanic	7	148	4.7
	Non-Hispanic	33	4,617	0.7

6.7.2 Computation of Missing Transcript Adjustments

The student transcript nonresponse adjustment factor for the h-th adjustment class was computed as follows:

$$STWAF_h = \frac{\sum_{ijk \in G(h)} w_{ijk} SCNRAF_a I[ij \in SR(\alpha)]\ TRIM(ij)\ STPSAF_g I[ijk \in E(g)]}{\sum_{ijk \in GR(h)} w_{ijk} SCNRAF_a I[ij \in SR(\alpha)]\ TRIM(ij)\ STPSAF_g I[ijk \in E(g)]} \qquad \text{(Equation 6.7.1)}$$

The set G(h) includes all graduated students in the h-th adjustment class, with the set GR(h) containing the subset of these students with complete and usable transcripts. The first factor in each term of each summation is the student base weight, discussed in Section 6.4.1. The second and third terms comprise the school nonresponse adjustment, discussed in Section 6.5.4. The fourth term is the school's trimming factor, discussed in Section 6.6.2, and the fifth and sixth terms are the student poststratification factors, discussed in Section 6.6.4.

These adjustments for missing transcripts are also necessary for the assessment linked weights and the excluded student linked weights. The same nonresponse cells were used as were used for the HSTS weights. The adjustments for each assessment a link weight are as follows:

$$STWFL_{ah} = \frac{\sum_{ijk \in G(h)} w^a_{ijk} SCNRFL_{aa} I[ij \in SRL_a(\alpha)]\ TRIM_a(ij)\ STNNRF_{a\gamma} I[ijk \in STR_a(\gamma)] STPSFL_{ag} I[ijk \in E(g)]}{\sum_{ijk \in GR(h)} w^a_{ijk} SCNRFL_{aa} I[ij \in SRL_a(\alpha)]\ TRIM_a(ij)\ STNNRF_{a\gamma} I[ijk \in STR_a(\gamma)] STPSFL_{ag} I[ijk \in E(g)]}$$

(Equation 6.7.2)

The first factor in each term in each summation is the assessment *a* student base weight, discussed in Section 6.4.3. The second and third factors comprise the school nonresponse adjustment factor for assessment weights, discussed in Section 6.5.5. The fourth factor is the assessment weight school trimming factor, discussed in Section 6.6.3. The fifth and sixth factors comprise the student assessment nonresponse adjustment, discussed in Section 6.6.1, and the remaining two factors are the student poststratification factor for the assessment weights, discussed in Section 6.6.5.

The corresponding missing transcripts adjustment for the excluded student weights was computed as follows:

$$STWFL_{eh} = \frac{\sum_{ijk \in G(h)} w^e_{ijk} SCNRFL_{a} I[ij \in SRL(\alpha)]\ TRIM_S(ij)\ STPSFL_{eg} I[ijk \in E(g)]}{\sum_{ijk \in GR(h)} w^e_{ijk} SCNRFL_{a} I[ij \in SRL(\alpha)]\ TRIM_S(ij)\ STPSFL_{eg} I[ijk \in E(g)]}$$ (Equation 6.7.3)

The first factor is the excluded student base weight, discussed in Section 6.4.3. The second and third factors are the school nonresponse adjustment, discussed in Section 6.5.5; the fourth factor is the "special" school trimming factor for excluded students, discussed in Section 6.6.3; the fifth and sixth factors are the student poststratification adjustments for excluded students, discussed in Section 6.6.6.

Table 6-23 presents the final nonresponse adjustment factors for the HSTS weights, each assessment linked weight, and the excluded student linked weight. The 11 nonresponse cells were collapsed into 4 cells for the excluded students because of small sample count. The adjustment given in the table is the overall adjustment for the larger cell.

Table 6-23. Nonresponse adjustment factors for missing transcripts

Cell Number	$STWAF_h$	Reading assessment $STWFL_{ah}$	History assessment $STWFL_{ah}$	Geography assessment $STWFL_{ah}$	Excluded students $STWFL_{ah}$
1	1.033	1.028	1.027	1.011	1.169
2	1.010	1.012	1.006	1.005	1.169
3	1.026	1.033	1.019	1.028	1.169
4	1.044	1.020	1.022	1.027	1.485
5	1.021	1.013	1.012	1.016	1.485
6	1.156	1.050	1.034	1.012	1.242
7	1.017	1.029	1.012	1.011	1.242
8	1.141	1.083	1.061	1.035	1.242
9	1.018	1.021	1.002	1.014	1.242
10	1.039	1.044	1.016	1.012	1.234
11	2.097	1.000	1.000	1.000	1.234

6.7.3 Final Sampling Weights

Final HSTS sampling weights were assigned to students in the HSTS study for which a transcript was received. These sampling weights are computed as follows:

$$W_{ijk} = w_{ijk}\, SCNRAF_{\alpha} I[ij \in SR(\alpha)]\, TRIM(ij)\, STPSAF_{g} I[ijk \in E(g)] * STWAF_{h} I[ijk \in GR(h)] \quad \text{(Equation 6.7.4)}$$

The first factor is the student base weight, discussed in Section 6.4.1. The second and third factors comprise the school nonresponse adjustment, discussed in Section 6.5.4. The fourth factor is the school's trimming factor, discussed in Section 6.6.2. The fifth and sixth factors comprise the student poststratification factors, discussed in Section 6.6.4. Finally, the remaining two factors comprise the student missing transcript adjustment factor, discussed in Section 6.7.2.

Final linked sampling weights were assigned to all students in the HSTS study for which transcripts were received and who were assessed using one of the NAEP assessments. These weights are computed for each assessment a as follows:

$$W^{a}_{ijk} = w^{a}_{ijk}\, SCNRFL_{a\alpha} I[ij \in SRL_a(\alpha)]\, TRIM_a(ij)\, STNNRF_{a\gamma} I[ijk \in STR_a(\gamma)] * STPSFL_{ag} I[ijk \in E(g)]\, STWFL_{ah} I[ijk \in GR(h)] \quad \text{(Equation 6.7.5)}$$

The first factor is the assessment *a* student base weight, discussed in Section 6.4.3. The second and third factors comprise the school nonresponse adjustment factor for assessment weights, discussed in Section 6.5.5. The fourth factor is the assessment weight school trimming factor, discussed in Section 6.6.3. The fifth and sixth factors comprise the student assessment nonresponse adjustment, discussed in Section 6.6.1. The seventh and eighth factors comprise the student poststratification factor for assessment weights, discussed in Section 6.6.5, and the final two factors are the missing transcripts adjustment factor for assessed weights, discussed in Section 6.7.2.

Final sampling weights were also computed for students in the HSTS study excluded from NAEP assessments, for which transcripts were also received. These weights are computed as follows:

$$W^e_{ijk} = w^a_{ijk}\, SCNRFL_\alpha I[ij \in SRL(\alpha)]\, TRIM_S(ij)\, STPSFL_{eg} I[ijk \in E(g)]^* \\ STWFL_{eh} I[ijk \in GR(h)] \qquad \text{(Equation 6.7.6)}$$

The first factor is the excluded student base weight, discussed in Section 6.4.3. The second and third factors are the school nonresponse adjustment, discussed in Section 6.5.5; the fourth factor is the "special" school trimming factor for excluded students, discussed in Section 6.6.3; the fifth and sixth factors are the student poststratification adjustments for excluded students, discussed in Section 6.6.6. The final factors are the student missing transcript adjustment for excluded students, discussed in Section 6.7.2.

Table 6-24 presents the distributions of these final weights for the HSTS weights (W_{ijk}), for the assessment linked weights for reading, history, and geography, respectively (W^a_{ijk}), and for excluded students (W^e_{ijk}). The tables include the count of students who have nonzero values of these weights, the total sum over all students of the weights, the minimum and maximum nonzero weights, and the quartiles for these weights.

Table 6-24. Distributions of the final HSTS and linked weights

Sample Distribution	HSTS weights	Reading assessment linked weights	History assessment linked weights	Geography assessment linked weights	Excluded student linked weights
Students with nonzero weights	25,335	9,258	5,070	4,143	533.00
Total (in thousands)	3,010	2,981	2,925	2,941	83.50
Minimum	1.72	30.94	21.80	13.06	25.93
25th percentile	64.65	182.96	328.03	396.49	101.70
Median	90.64	274.38	490.79	608.26	144.26
75th percentile	157.78	433.52	778.98	953.64	204.24
Maximum	829.29	3,216.7	2,021.8	2,751.5	349.07

6.7.4 Final Replicate Weights

The computation of final replicate school base weights is discussed in Section 6.4.5. It is only for this component that the replicate weights differ. The remaining weights and adjustments are computed as they were for the primary weights. The HSTS student base weights and student linked base weights are computed as follows:

$$w_{ijk}(r) = w_{ij}(r)w_{k|ij}$$
$$w^a_{ijk}(r) = w_{ijk}(r)w_{a|ij}w_{a|ijk}I[ijk \in U_a] \qquad \text{(Equation 6.7.7)}$$
$$w^e_{ijk}(r) = w_{ijk}(r)I[ijk \in U_e]$$

These quantities are defined in Sections 6.4.3 and 6.4.5. Note that all of these base weights are identical to the corresponding "main" (nonreplicate) base weights except for the factor $w_{ij}(r)/w_{ij}$.

In principle, the replicate weights should repeat the entire process of computing the final weights using the new replicate base weights. This replication will capture any components of variability introduced to the final weights by these processes. This was done for the HSTS and linked weights for most of these processes, except for the trimming step preceding poststratification, and the two CHAID analyses which selected school and missing transcript nonresponse cells.

The same trimming factors and CHAID categories were used for calculating the replicate weights as for the main weights. The components of variability introduced by these processes should be relatively small, so the complexity of replicating these processes led us to forgo replication of these

processes along with the basic nonresponse and poststratification steps. We note that the trimming process was also not replicated in the development of the 1994 NAEP replicate weights.

For the school nonresponse adjustments then the same nonresponse cells were used as for the "main" weight school nonresponse adjustments (these cells are presented in Table 6-4). The nonresponse adjustments were all recomputed for each replicate weight using the new replicate school base weights:

$$SCNRAF_{\alpha}(r) = \frac{\sum_{ij \in S(\alpha)} w_{ij}(r) G_{ij}}{\sum_{ij \in SR(\alpha)} w_{ij}(r) G_{ij}} \qquad \alpha = 1,...,9, \;\; r = 1,...,62. \qquad \text{(Equation 6.7.8)}^{22}$$

The quantities $S(\alpha)$, $SR(\alpha)$, and G_{ij} are defined in Section 6.5.4. The corresponding replicate weights $SCNRFL_{\alpha}(r)$ and $SCNRFL_{a\alpha}(r)$ are defined in a similar manner: replacing w_{ij} with $w_{ij}(r)$ in Equations 6.5.6 and 6.5.7, respectively.

The replicate student nonresponse adjustments are based on the same set of cells as were used for the main student nonresponse adjustments $STNNRF_{a\gamma}$ and $STNNRF_{\gamma}$ (see Section 6.6.1). These replicate adjustments for the assessment groups were computed as follows:

$$STNNRF_{a\gamma}(r) = \frac{\sum_{ijk \in ST_a(\gamma)} w^a_{ijk}(r) SCNRFL_{a\alpha}(r) I[ij \in SRL_a(\alpha)]}{\sum_{ijk \in STR_a(\gamma)} w^a_{ijk}(r) SCNRFL_{a\alpha}(r) I[ij \in SRL_a(\alpha)]} \qquad \text{(Equation 6.7.9)}$$

This equation is analogous to Equation 6.6.1. A corresponding definition for $STNNRF_{\gamma}(r)$ can be generated modifying Equation 6.6.2 in a similar manner.

The poststratification adjustments were also replicated, using the same poststratification cells and poststratification control totals as were used for the main weights. The replicate poststratification adjustment for the HSTS weights is defined as follows:

---

[22] Actually r = 1,...,26,28,...,62 as explained in Section 6.8.

$$STPSAF_g(r) = \frac{C_g}{\sum_{ijk \in E(g)} w_{ijk}(r) SCNRAF_\alpha(r) I[ij \in SR(\alpha)] TRIM(ij)} \qquad \text{(Equation 6.7.10)}$$

This equation is analogous to Equation 6.6.16. Note that the trimming factor is from the main weights analysis -- that is, it is not replicated, also. Similar modifications of Equations 6.6.17 and 6.6.18 define replicate adjustments $STPSFL_{ag}(r)$ and $STPSFL_{eg}(r)$.

The final step in computing the final replicate school weights was to replicate the missing transcript adjustments. The missing transcript adjustment cells were the same as were used for the main weights (as given in Table 6-21). Following Equation 6.7.1 the replicate missing transcript adjustment factor for the HSTS weights is given as:

$$STWAF_h(r) = \frac{\sum_{ijk \in G(h)} w_{ijk}(r) SCNRAF_\alpha(r) I[ij \in SR(\alpha)]\ TRIM(ij)\ STPSAF_g(r) I[ijk \in E(g)]}{\sum_{ijk \in GR(h)} w_{ijk}(r) SCNRAF_\alpha(r) I[ij \in SR(\alpha)]\ TRIM(ij)\ STPSAF_g(r) I[ijk \in E(g)]}$$

(Equation 6.7.11)

Similar modifications of Equations 6.7.2 and 6.7.3 give us the replicate adjustments $STWFL_{ah}(r)$ and $STWFL_{eh}(r)$.

The final replicate weights used in any jackknife variance calculation were computed as follows (analogous to Equations 6.7.4, 6.7.5, and 6.7.6):

$$W_{ijk}(r) = w_{ijk}(r)\ SCNRAF_\alpha(r) I[ij \in SR(\alpha)]\ TRIM(ij)\ STPSAF_g(r) I[ijk \in E(g)] *$$
$$STWAF_h(r) I[ijk \in GR(h)]$$

(Equation 6.7.12)

$$W^a_{ijk}(r) = w^a_{ijk}(r)\ SCNRFL_{a\alpha}(r) I[ij \in SRL_a(\alpha)]\ TRIM_a(ij)\ STNNRF_{a\gamma}(r) I[ijk \in STR_a(\gamma)] *$$
$$STPSFL_{ag}(r) I[ijk \in E(g)]\ STWFL_{ah}(r) I[ijk \in GR(h)]$$

(Equation 6.7.13)

$$W^e_{ijk}(r) = w^e_{ijk}(r)\ SCNRFL_\alpha(r) I[ij \in SRL(\alpha)]\ TRIM_S(ij)\ STPSFL_{eg}(r) I[ijk \in E(g)] *$$
$$STWFL_{eh}(r) I[ijk \in GR(h)]$$

(Equation 6.7.14)

7. 1994 HIGH SCHOOL TRANSCRIPT STUDY DATA FILES

Data from the 1994 High School Transcript Study are organized into eight data files encompassing the different levels of information: (1) Master CSSC File; (2) Course Offerings File; (3) School File; (4) Student File; (5) Linked Weights File; (6) IEP/LEP Questionnaire File; (7) Tests and Honors File; and (8) Transcript File. The relationships among the files are shown in Figure 7-1. Except for the Master CSSC File (which is not related to individual schools or students), all files can be linked by PSU and school identifiers. The Student, IEP/LEP Questionnaire, Transcript, Linked Weights, and Tests and Honors Files can be linked by student identifiers; and the Master CSSC can be linked to the Course Offerings or Transcript File by CSSC number.[23]

To identify a specific school, the PSU and school IDs must be used in combination. Each school has a unique PSU/School ID combination and all student IDs are unique. For students in the 280 schools that are fully linked to NAEP, student IDs are their 10-digit NAEP booklet numbers. All other students were assigned unique 10-digit IDs beginning with 990.

Weights, developed using the procedures described in Chapter 6, are contained in the Student File and the Linked Weights File. We have provided the final student weight (FINSTUWT) in the Student File and the final linked weight (FINLNKWT) in the Linked Weights File so that data analyses can be weighted up to national totals. The final student weight should be used in analyses involving only transcript data. The weights in the Linked Weights File should be used in analyses involving both transcript data and data obtained from NAEP data files.

7.1 Master CSSC File

The Master CSSC File contains all codes in the modified version of the *Classification of Secondary School Courses* (CSSC) used in this study. There are 2,185 records, sorted by CSSC number. In addition to the original six-digit CSSC codes created in 1982, the file contains the codes added for the 1987 and 1990 studies and 12 additional codes added during the current study.

---

[23] This chapter provides a short description of the 1994 HSTS files. For a full description, see Legum et al (1991). *The 1994 High School Transcript Study Data File User's Manual*, Washington, DC, U.S. Department of Education, Office of Educational Research and Improvement, NCES 97-025.

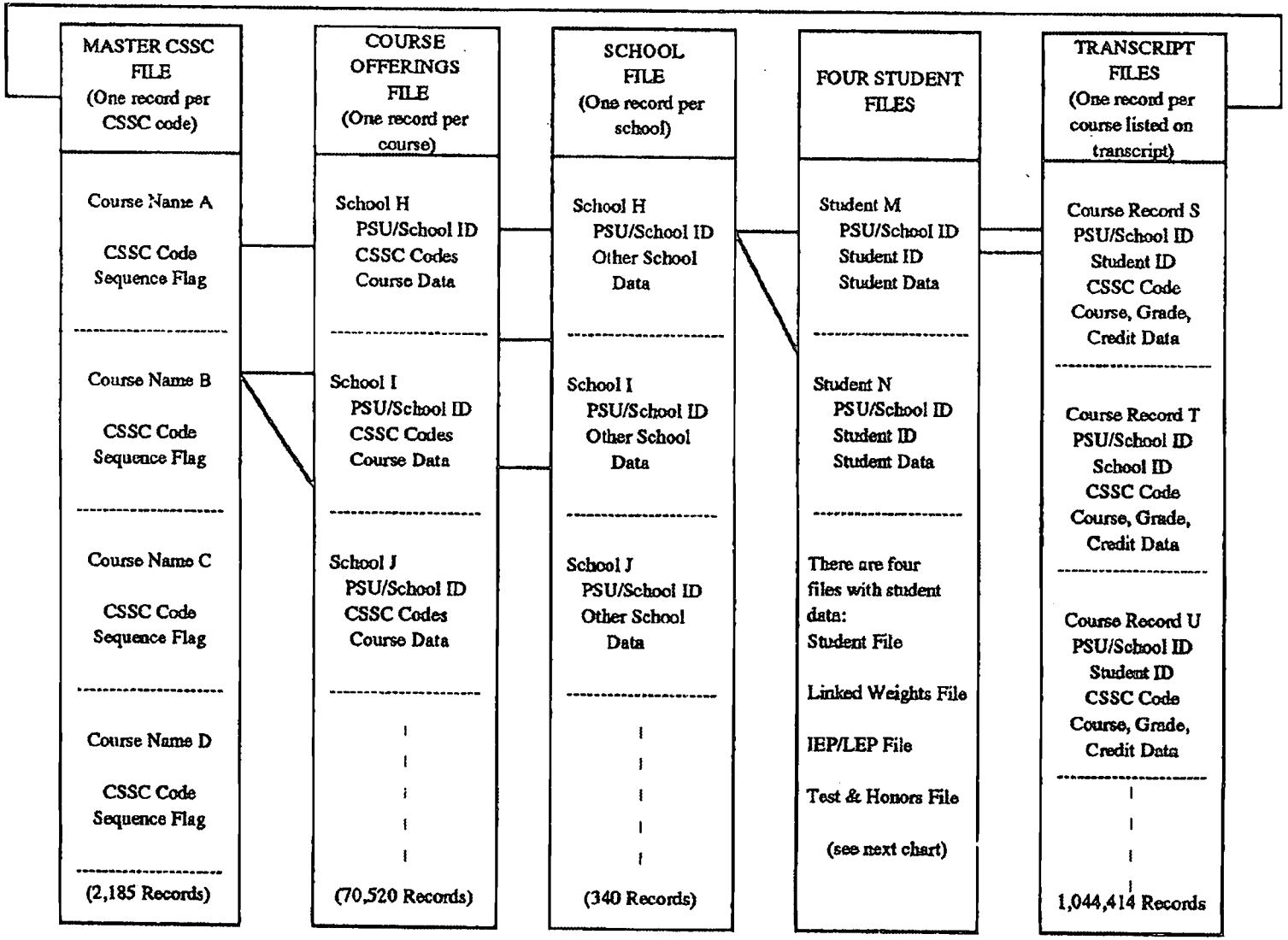
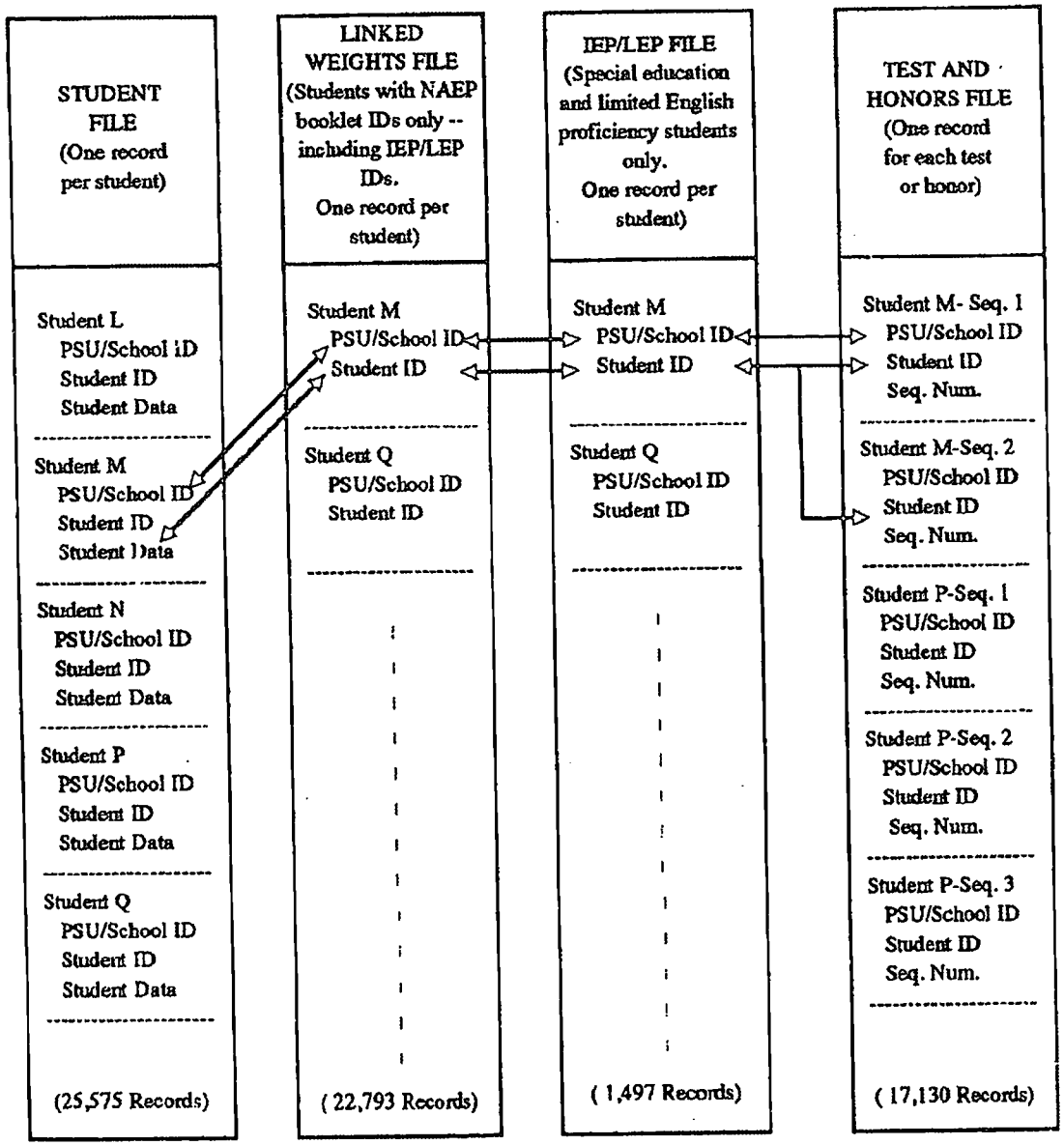


Figure 7-1. Project data files and linking identifiers



All student files can be linked using the PSU, School ID, and Student ID variables.

Note that the STUDENT file contains records for all students. The other three files contain records for subsets of the students in the STUDENT file.

Figure 7-1. Project data files and linking identifiers (continued)

The new codes are documented in Appendix E, *1994 Additions to the Classification of Secondary School Courses*. These codes were added when courses were encountered on the transcripts that were clearly different from codes already contained in the CSSC. No new two-digit or four-digit categories were added during the 1994 transcript study.

A special education flag (SPEDFLAG), an expansion to the CSSC initiated during the 1987 transcript study, was retained as part of the current version of the CSSC. When a course on a transcript was limited in enrollment to special education students, it was coded using the regular CSSC code with a special education indicator of "0" or "2".[24] Any course not so limited has the special education flag set to "1".

As in the 1990 transcript study, all CSSC entries have been coded with a sequence flag. A "0" value for the sequence flag indicates that the course is not part of an instructional sequence. A "1" indicates that the course is the first course in an instructional sequence, and a "2" indicates that the course is an advanced course in an instructional sequence (i.e., not the initial course in the sequence). The CSSC Master File is organized by the CSSC code and contains four variables: the CSSC course code, the special education flag, the sequence flag, and the standard course title.

7.2 Course Offerings File

The Course Offerings File is organized by school and contains one record for each course listed in the school's course catalog or appearing on a student's transcript as a non-transfer course taken at that school. Each of the 70,520 records contains the PSU, school ID, course title, course CSSC code, special education flag, the source of the catalog (e.g., generated from transcripts or from a school-provided catalog) and six additional pieces of information about the course: (1) the location of the course (including various off-campus locations); (2) the language of instruction; (3) whether or not it was remedial or below-grade-level course; (4) whether or not it was an honors-level course; (5) if it was a combination course (i.e., composed of more than one part, requiring more than one CSSC code for accurate description); (6) if it was part of an instructional sequence. The file is sorted by the PSU and school ID numbers.

[24] The values of the SPEDFLAG variable are as follows: 0 = a functional level course limited in enrollment to special education students; 1 = a regular course not limited in enrollment to special education students; 2 = a special education course not at the functional level, but limited in enrollment to special education students.

The Course Offerings File is a complete listing of courses offered in all participating schools that provided us with school-level course catalogs It contains all courses listed in the school-level course catalogs received and any non-transfer courses listed on the transcripts not otherwise appearing in the catalogs. For example, in a school with grades 10 through 12 whose students all take 9th grade in a junior high, the 9th-grade courses are not treated as transfer courses, but appear as if they were offered by the high school. This treatment provides a more balanced picture of the courses available to American students in 4 years of high school than would be provided by treating such courses as transfer courses. For the 22 schools from which we did not receive a catalog, the list of unique course titles appearing on the sampled transcripts is the only available source of course offering entries. A complete listing of all courses included on the transcripts can be extracted only from the Transcript File, since transfer courses do not appear in the Course Offerings File.

7.3 School File

The School File is sorted by PSU and school ID and contains one record for each of the 340 participating schools. School variables gathered during the transcript study are included, as well as the school's responses to the NAEP School Characteristics and Policy Questionnaire (see Appendix B).

7.4 Student File

The Student File contains one record for each of the 25,575 graduates who were identified. Since 81 transcripts were not received, full transcript information is included for the 25,494 graduated students for whom transcripts were obtained and coded.[25] Students are identified by PSU, School, and Student ID variables, and the file is sorted by this group of variables The file contains the demographic information gathered for each student, sampling information, weights to be used in analysis, and replicate weights for variance estimation. The final student weight for each student is the variable FINSTUWT. The component weights used to derive the final student weight are also included. In addition, the file contains a flag indicating whether or not the student is disabled

[25]For the 81 students, we usually know their gender, race, birth year, birth month, whether they had an IEP, whether they were classified as LEP, and whether they received Chapter I services.

and a condition variable indicating the specific nature of the disability when applicable.[26] The file also contains a series of derived variables including one designating the student's academic track as academic, vocational, both, or neither, and summaries of the student's course-taking record by major educational topic.

Note that 211 students have final student weights (FINSTUWT) of zero. Of these, 81 are the students for whom we obtained no transcripts. There are 110 students receiving regular or honors diplomas (EXSTAT=1 or 2) whose transcripts do not have enough codable courses to account for at least 75 percent of the Carnegie units required by their schools to graduate (i.e., GRREQFLG=4). They were given final weights of zero. In other words, only transcripts fully documenting at least 3 years of high school received positive weights. There are 20 students with a GRREQFLG value of 4 who were given positive weights. Nine of these received special education diplomas and 11 received certificates of attendance. Their transcripts fully documented at least 3 years of high school even though the total number of credits is less than 75 percent of the total required for a regular diploma.

The weights included on the student file are for all students in the study, both those we can link to the NAEP assessment and those we cannot. Analyses of just the linked students must take into account a different set of nonresponse adjustments than the unlinked weights (see Chapter 6). The appropriate weights to be used in such a linked analysis are contained in the Linked Weights File.

7.5 Linked Weights File

The Linked Weights File contains the set of weights needed to perform analyses on the subset of schools and students fully linked to the NAEP assessment. As discussed in Chapter 6, because different sets of schools were eligible to participate in the NAEP and the HSTS studies, and because different sets of schools chose to participate in each, different school-level nonresponse adjustments need to be used when constructing student weights. For similar reasons, different student-level nonresponse adjustments need to be used when constructing student weights. Furthermore, since the main 1994 NAEP study consisted of three parallel sets of assessments (reading, geography, and history), separate sets of weights need to be used for each assessment. In addition, we have provided a

[26] The values of the disabling condition code are 00-not disabled, 01-multiple disabilities, 02-mentally retarded, 03-hard of hearing, 04-deaf, 05-speech-impaired, 06-visually impaired/blind, 07-deaf/blind, 08-emotionally disturbed, 09-orthopedically impaired, 10-learning disabled, 11-other disability, and 99-not ascertained.

separate set of weights for students who were excluded from the NAEP assessments on the basis of a disability or limited English proficiency.

The Linked Weights File contains one record for each of the 22,793 graduates for whom we have NAEP booklet numbers. As in the Student File, students are identified by the combination of PSU, School, and Student ID variables. The file is sorted by these identifier variables. The first three digits of the student ID identify the assessment in which the student participated. Values between 001-022 indicate reading; 031-049, geography, and 101-133, history.[27] For ease of use, this file also contains the demographic variables included on the Student File. The final linked weight variable is FINLNKWT.

7.6 IEP/LEP Questionnaire File

School special education staff members were asked to fill out an IEP/LEP Questionnaire for each disabled student and each student with limited English proficiency who was sampled for the study. The IEP/LEP Questionnaire File contains one record for each of 1,497 students, with data from these completed questionnaires. The file is sorted by PSU, School, and Student ID.

7.7 Tests and Honors File

The Tests and Honors File contains information on standardized test scores and honors that appear on high school transcripts. Of the transcripts collected, 6,636 (26.0 percent) contained either standardized test scores or notations regarding honors and awards that students received. The Tests and Honors File lists this information. Because of the relatively small percentage of the transcripts represented, the data in this file should be used with caution.

As in the Student File, students are identified by the combination of PSU, School, and Student ID variables. The file is sorted by these identifier variables. Each entry on a transcript is identified with a unique sequence number (unique within student). Entries are sorted by sequence number within student. Each entry also contains an indicator of the record type ("T" = test,

[27] One other set of student ID prefixes appears on the Student File, but not on the Linked Weights File. The prefix "990" is used for all non-linked students-- that is, students in schools for whom a sample was drawn in the field for the transcript study.

"H" = honor), the month and year of the test or honor (if available), the semester (Fall or Spring, if available), and a 40 character description of the honor or the test. For most tests, we have also provided the test score. Although it was not always possible to provide meaningful entries for some test scores (e.g., some schools reported SRA tests with percentiles and some with scaled scores) and the subtests which are reported varied tremendously, we provide complete scores for the PSAT math and verbal subtests, the SAT math and verbal subtests, and five ACT subtests. The remaining test information is of interest in so far as it can be used to determine the distribution of test data being reported on high school transcripts. The file contains 17,130 records.

7.8 Transcript File

The Transcript File contains one record for each course appearing on the sampled students' transcripts. This is an extremely large file, containing 1,044,441 records. Courses are identified by PSU, School, Student ID, and course sequence number (within students). The records in the file are sorted by PSU, school, student ID, and course sequence number. Variables for each course record include grade level when taken, school year when taken, course title, grade received (original and standardized), credit received (original and standardized), course CSSC code, if taught off campus, if taught in a language other than English, if it is a remedial or below-grade-level course, and if it is an honors course.

7.9 NAEP Data Files

There are three NAEP data files containing proficiency scores for each student who completed the assessment. These files are the 1994 NAEP Reading Data File; the 1994 NAEP Geography Data File; and the 1994 NAEP History Data File.

These files contain the NAEP scores for 1994 graduates who participated in a NAEP assessment in a school that is fully linked to the High School Transcript Study. In the case of the Geography and History scores, these files contain scores for all graduates who participated in NAEP. In the case of the Reading scores, these files contain scores for all graduates who participated in the NAEP Reading assessment, but do not contain scores for a large number of graduates who were part of a special psychometric study that did not provide comparable scores.

Because NAEP scores are designed to provide accurate group estimates rather than student-level information, they are "conditioned" on other variables (e.g. Parents' education level and NAEP region) in the NAEP datasets to provide more unbiased estimates when NAEP data are analyzed in conjunction with the conditioning variables.[28] The conditioning process has the effect of increasing the bias when analyses are made between NAEP scores and variables not in the conditioning set. In order to make the transcript data as usable as possible, Westat asked the Educational Testing Service to add transcript study variables to the conditioning process. The following variables were included in this analysis:

- ACAD_TRK — Student Program
- CLRANK/CLSIZE — Class Rank divided by Class Size
- EXSTAT — Student Exit Status
- GPA_C — Calculated GPA
- GRREQFLG — Graduation Requirements Level Flag
- HCFLAG — Student Disability Status
- REGION — Census Region
- STUB0100 - STUB1600 — These "stub" variables represent the number of credits students received in various subject areas. These are defined in detail in Appendix D of the Data File User's Manual.
- STUB2001 - STUB2005 — New Basics Curriculum categories. These variables represent variants of academically oriented course-taking patterns described in the *Nation at Risk* report. They are defined in detail in Appendix D of the Data File User's Manual.

All of the variables normally used by Educational Testing Service for conditioning the NAEP scores were also considered in the conditioning process. Thus all the variables in the transcript study Student File can be safely used in analyses involving NAEP scores. Because additional variables were included in the conditioning of NAEP scores for the transcript study, the NAEP scores reported

[28] See the NAEP 1994 Technical Report for a detailed discussion of conditioning.

in these files are slightly different from those contained in the records for the same students distributed solely as NAEP data.

Because fewer schools and students participated in both NAEP and HSTS than in either study alone, a different set of nonresponse adjustments applies to analyses using variables from both studies than for analyses confined to a single study. The weights in the Linked Weights File should be used in analyses comparing the NAEP data to the transcript data rather than the weights contained in the Student File. Note that if we do not have a complete transcript for a student, his or her weight is set to zero in the Linked Weights File.

The PSU, School, and Student IDs in the NAEP data files have the same structure as the corresponding variables in other transcript study files. If the need arises to match transcript study records with records obtained from NAEP files obtained from other sources, the analyst needs to be aware of the following differences in naming conventions:

Transcript study record identifier		NAEP record identifier (other than those distributed with the transcript files)	
Variable Name	Field Length	Variable Name	Field Length
PSU	3	PSU	3
SCHOOL	3	SCH	3
STUDENT	10	BOOK BKSER CHKDIG	3 6 1

The student identifier, STUDENT, in the transcript study is created by concatenating the NAEP book number (BOOK, which identifies the form of the assessment which was administered), the book serial number (BKSER), and the check digit (CHKDIG). The values of STUDENT are sufficient to uniquely identify a student in either the 1994 HSTS files or the 1994 NAEP files.[29]

[29]For students not linked to NAEP, the first 3 digits of the variable STUDENT are "990." The next 4 digits are a unique school identifier generated solely to ensure that the student identifiers are unique. The last 3 digits were sequentially assigned, starting with 001, to students within a school.

The following table summarizes the number of records in each NAEP data file and the corresponding number of non-zero weights in the Linked Weight File.

NAEP Data File	Number of records	Number of non-zero weights
Reading	6,502	6,475[30]
Geography	4,159	4,143
History	5,081	5,070

[30] There are 9,258 non-zero weights for students taking reading assessments; 2,783 of these students completed test versions of the reading assessment. Since their results were not conditioned, their data do not appear in the NAEP Reading File.

8. REFERENCES

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: John Wiley & Sons.

Johnson, E.G. and Allen, N.L. (1992). *The NAEP 1990 Technical Report.* Princeton, NJ: Educational Testing Service.

Jones, C., Clarke, M., Mooney, G., McWilliams, H., Crawford, I., Stephenson, B., and Tourangeau, R. (1983a). *High School and Beyond First Follow-Up (1982): Technical Report.* Chicago, IL. National Opinion Research Center, June 1983. (ERIC #ED 249-244).

Jones, C., Knight, S., Butz, M., Crawford, J. and Stephenson, B. (1983b). *High School and Beyond Transcripts Survey (1982): Data File User's Manual.* Washington, DC, U.S. Department of Education, National Center for Education Statistics, NCES 84-205, July 1983. (ERIC #ED243957).

Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 119-127.

Kish, L. and Frankel, M. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Series B*, 36, 1-22.

Krewski, D. and Rao, J. (1981). Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication. *Annals of Statistics*, 9, 1010-1019.

Legum, S., Caldwell, N., Goksel, H., Haynes, J., Hynson, C., Rust, K., and Blecher, N. (1993). *The 1990 High School Transcript Study Technical Report.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement NCES 93-117.

Little, R. J. A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley & Sons.

Ludwig, M.J., Olivetti, L., Sandberg, N., and Waite, B. (1982). *A Classification of Secondary School Courses.* Alexandria, VA: Evaluation Technologies, Inc.

Morgan, R., Hunt, E., and Carpenter, J. (1991). *Classification of Instructional Programs.* NCES 91-396. U.S. Department of Education, Office of Educational Research and Improvement.

National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform.* Washington, DC. (ERIC No. ED 226 006)

Oh, H.L. and Scheuren, F.S. (1983). Weighting adjustments for unit nonresponse. In W.G. Madow, I. Olkin, and D.B. Rubin (Eds.), *Incomplete Data in Sample Surveys, Vol. II: Theory and Annotated Bibliography.* New York: Academic Press.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics*, 1(4), 381-397.

Rust, K., Bethel, J., Burke, J., and Hansen, M. (1990). *The 1988 National Assessment of Educational Progress - Sample and Weighting Procedures, Final Report*. Rockville, MD: Westat, Inc.

Thorne, J., Burke, J., Rust, K., Marshall, R., Caldwell, N., Sickles, D., and Ha, P. (1989). *1987 High School Transcript Study Technical Report*. Rockville, MD: Westat, Inc.

Wallace, L. and Rust, K. (1995). *1994 National Assessment of Educational Progress Sampling and Weighting Procedures, Final Report*.

Wolter, K. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

APPENDIX A

STUDY MATERIALS SENT TO SCHOOLS AND DISTRICTS

U.S. DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

NATIONAL CENTER FOR EDUCATION STATISTICS

January 1994

Dear Superintendent:

As described in previous mailings to your district, the 1994 High School Transcript Study is being conducted in conjunction with the 1994 National Assessment of Educational Progress (NAEP). The purpose of this study is to supply data to educational researchers and policy analysts on course-taking patterns and to examine the relationship of these patterns to achievement in secondary schools. The NAEP school sample is being used both because it is a nationally representative sample and in order that NAEP data and transcript data can be linked for schools participating in both. The participation of all selected schools (regardless of whether they are participating in NAEP) is needed to make the results of the transcript study comprehensive, accurate, and timely.

A list of the NAEP schools in your district selected for this study is enclosed. Detailed information on transcript activities in the school accompanies this letter. **No student time is involved.** Students' names and other individually identifying information will be removed from copies of the transcripts before they leave the school, and schools will be reimbursed at the standard rate for supplying transcripts.

Initial activities will be conducted at the same time NAEP supervisors are in the schools selecting the NAEP sample. In the fall of 1994, supervisors will return to the school to collect the requested transcripts.

The granting of Education Department authority for collection of the transcript data has been made pursuant to the provisions of the Family Education Rights and Privacy Act (FERPA) (20 U.S.C. 1232g), as implemented by 34 CFR 99.31(a)(3)(ii) and 99.35. These laws and regulations permit an educational agency to disclose records to authorized representatives of the Secretary of Education without the prior consent of the survey participants in connection with the audit and evaluation of Federal and State supported education programs. The privacy of the information schools are asked to supply to the NAEP contractors will be protected as required by FERPA and will be further protected by the removal of names and other identifying information. A copy of the relevant section of FERPA regulations is reproduced on the reverse side of this page.

I would appreciate your cooperation in this important component of the 1994 NAEP. If you have any questions about the study or its procedures, I may be contacted at the Department of Education or you may contact Nancy Caldwell of Westat, Inc., at (800) 283-6237.

Sincerely,

Steve Gorman

Steve Gorman
Project Officer

WASHINGTON, D.C. 20208—

THE NATION'S REPORT CARD

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

1650 RESEARCH BOULEVARD • ROCKVILLE, MARYLAND 20850
TELEPHONE 1-800-283-6237 • FAX 301-294-2038

January 1994

Dear Principal:

Thank you for your participation in the 1994 National Assessment of Educational Progress. As indicated in the letter from Steve Gorman of the National Center for Education Statistics and as described in previous informational mailings regarding the 1994 national assessment, the U.S. Department of Education has authorized the National Assessment of Educational Progress (NAEP) to collect high school transcript data.

The purpose of this study is to obtain current information on course-taking patterns of high school students and to correlate this information with achievement data from the 1994 NAEP. To be nationally representative, the 1994 High School Transcript Study will include a sample of secondary schools selected for the 1994 National Assessment of Educational Progress. This is an important study and each participating school will make a valuable contribution to its success.

Detailed information on transcript activities in the school accompanies this letter. The activities for Phase 1 will be conducted at the same time that NAEP supervisors are in your school selecting the NAEP sample. Phase 2 of the study will occur in the fall of 1994 when the NAEP supervisor will return to your school to collect the requested transcripts. No student time is involved and schools will be reimbursed at the standard rate for supplying transcripts.

NAEP has been authorized to collect information on sampled students from their academic records pursuant to the provisions of the Family Education Rights and Privacy Act (FERPA). All students' names and other individually identifying information will be removed from the collected data before it is sent to our offices. All information obtained through this study will be kept confidential and will only be used for statistical reporting purposes.

Should you have any questions, please contact either me or Sandra Rieder at Westat (800) 283-6237.

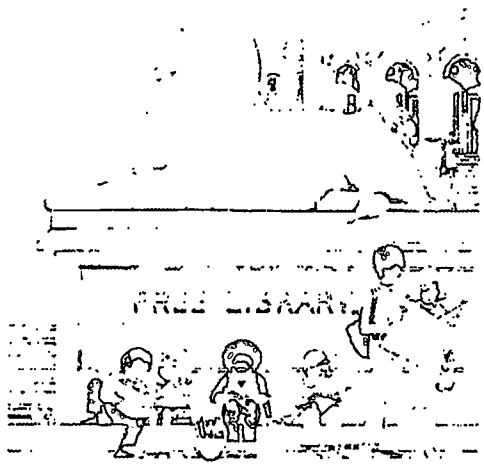
Sincerely,

Nancy W. Caldwell

Nancy W. Caldwell
NAEP Project Director

APPENDIX B

1993-94 SCHOOL CHARACTERISTICS AND POLICIES QUESTIONNAIRE

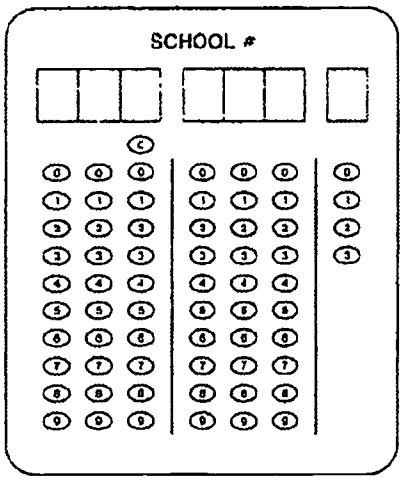


93 - 94
Grade 12

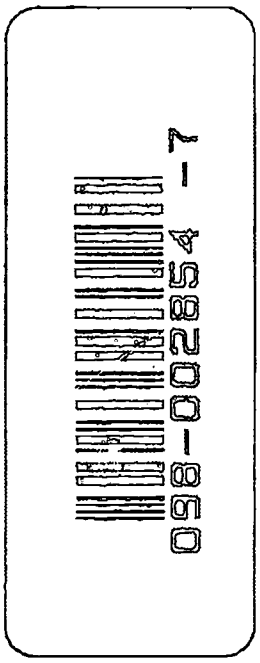
School Characteristics and Policies Questionnaire

Q-098

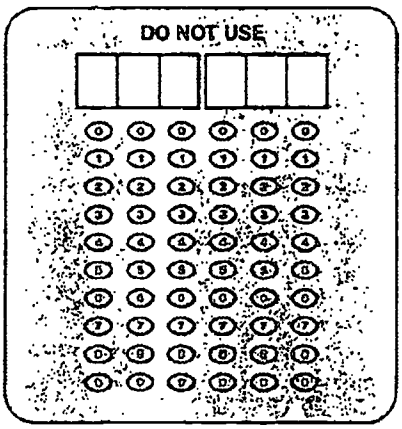
SCHOOL #



THE NATION'S REPORT CARD



DO NOT USE



A project of the Office of Educational Research and Improvement.
This report is authorized by law (20 USC 1221 e-1(i)). While you are not required to respond, your cooperation is needed to make the results of this survey comprehensive, accurate, and timely.

OMB NO 1850-0628 • Approval Expires 8/94
Mark Reflex® by NCS EP-153875-001:321 Printed in U.S.A.

During the 1993-94 school year, a sample of students across the country, including some students from your school, will be given a series of questions as part of the National Assessment of Educational Progress (NAEP). The current assessment focuses on achievement in reading, world geography, and U.S. history. As part of the assessment, NAEP will investigate the relationship between students' achievement and various school, teacher, and home factors that may influence this achievement. We are asking your school to complete this questionnaire about school factors. **This questionnaire should be completed by the principal or other head administrator.**

We realize that you are very busy; however, we urge you to complete the questionnaire as carefully as possible. The information that you provide will be kept confidential.

NAEP is authorized under Public Law 100-297. While your participation is voluntary, your responses to these questions are needed to make this survey accurate and complete.

Please answer directly on the questionnaire by filling in the appropriate oval.

When you are finished, please return the questionnaire to your school's NAEP coordinator.

Thank you very much for your help.

School Characteristics and Policies Questionnaire

Please fill in one oval for each question. Questions 1 through 82 should be completed by the principal or the head of the school.

Questions 1 – 5. Are twelfth grade students typically assigned to classes by ability and/or achievement levels (so that some classes are higher in average ability and/or achievement levels than others) in any of the following subjects? Fill in **one** oval on each line.

	Yes	No	
1. English	(A)	(B)	C035001
2. Mathematics	(A)	(B)	C035002
3. Science	(A)	(B)	WP000091
4. History	(A)	(B)	WP000092
5. Geography	(A)	(B)	WP000093

Questions 6 – 13. Beginning with 9th grade, how many semesters (or equivalent) of course work does your school or district require of each student in each of the following subjects for graduation from 12th grade by June 1994? Fill in **one** oval on each line. HE000964

	None	1	2	3	4	5	6	7	8	
6. English/literature/writing	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	HE000965
7. Mathematics	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	HE000966
8. Science	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	HE000967
9. Computer science	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	HE000968
10. Social studies	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	LC000506
11. History	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	LC000507
12. Geography	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	LC000508
13. Foreign languages	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	HE000970

Questions 14 – 24. Are courses of at least one semester in length taught in your school in each of the following subjects? Fill in **one** oval on each line. LC000509

		Yes	No	
14.	Computer science	(A)	(B)	WP000094
15.	Calculus	(A)	(B)	LC000512
16.	World geography or other regional geography	(A)	(B)	LC000513
17.	Advanced biology	(A)	(B)	LC000514
18.	Advanced chemistry	(A)	(B)	LC000515
19.	Advanced physics	(A)	(B)	LC000516
20.	Trigonometry	(A)	(B)	WP000095
21.	Pre-calculus, third-year algebra, elementary functions, or analysis	(A)	(B)	WP000096
22.	Probability and/or statistics	(A)	(B)	WP000097
23.	Unified, integrated or sequential mathematics	(A)	(B)	WP000098
24.	U.S. History	(A)	(B)	WP000099

25. Is there a district or state test that students in your school are required to take at any of the following grades? Fill in as many ovals as apply, but only for grades taught at your school. C035401

(A) Not required at any grade

(B) Grade 9

(C) Grade 10

(D) Grade 11

(E) Grade 12

Questions 26 – 28. Are computers available to students in your classes in any of the following ways? Fill in **one** oval on each line.

		Yes	No	
26.	Available all the time in classrooms	A	B	C035701
27.	Grouped in a separate computer laboratory available to classes	A	B	C035702
28.	Available to bring to classrooms when needed	A	B	C035703

Questions 29 – 35. Of the students in your school, approximately what percentage receive the following services? Fill in **one** oval on each line.

		None	1-5%	6-10%	11-25%	26-50%	51-75%	76-90%	Over 90%	
29.	Subsidized school lunch and/or nutrition program	A	B	C	D	E	F	G	H	C032001
30.	Remedial reading instruction	A	B	C	D	E	F	G	H	C032002
31.	Remedial mathematics instruction	A	B	C	D	E	F	G	H	C032003
32.	Bilingual education	A	B	C	D	E	F	G	H	C032004
33.	English-as-a-second-language instruction (not in a bilingual education program)	A	B	C	D	E	F	G	H	C032005
34.	Special education for disabled students	A	B	C	D	E	F	G	H	C032006
35.	Gifted and talented education	A	B	C	D	E	F	G	H	LC000477

Questions 36 – 40. How many students in your school are currently enrolled in Advanced Placement courses in each of the following subjects? Fill in **one** oval on each line.

	None	1-5	6-10	11-25	26-50	More than 50	
36. English	A	B	C	D	E	F	C035802
37. Science	A	B	C	D	E	F	WP000100
38. History	A	B	C	D	E	F	WP000101
39. Geography	A	B	C	D	E	F	WP000102
40. Calculus	A	B	C	D	E	F	WP000103

Questions 41 – 45. Does your school do any of the following to involve parents? Fill in **one** oval on each line.

	Yes, routinely	Yes, occasionally	No	
41. Use parents as aides in classrooms	A	B	C	C032207
42. Encourage parents to visit classrooms	A	B	C	C032208
43. Have parents review or sign students' homework	A	B	C	LC000482
44. Assign homework for students to do with parents	A	B	C	LC000484
45. Have a parent volunteer program	A	B	C	LC000486

Questions 46- 47. For the next two questions, please write the appropriate percentages in the boxes. Please PRINT LEGIBLY. Use one character per block in the indicated areas. Keep all printing within the boxes. Do not make any stray marks. Use only a No. 2 pencil.

46. Of the students attending your school, approximately what percentage lives in each of the following areas?

LC000478

A In a rural area with a population of less than 2,500 ☐☐☐.0%

B In a town with a population of 2,500 to 9,999 ☐☐☐.0%

C In a town with a population of 10,000 or more ☐☐☐.0%

Ex: 50% would be written as 050.0%

Examples of numerals are: 1 2 3 4 5 6 7 8 9 0

47. Of the students attending your school, approximately what percentages are children of:

LC000479

A professional or managerial personnel ☐☐☐.0%

B sales, clerical, technical or skilled workers ☐☐☐.0%

C factory or other blue-collar workers ☐☐☐.0%

D farm workers ☐☐☐.0%

E persons not regularly employed but not on welfare ☐☐☐.0%

F persons on welfare ☐☐☐.0%

Questions 48 – 53. How would you characterize each of the following within your school? Fill in **one** oval on each line.

		Very positive	Somewhat positive	Somewhat negative	Very negative	
48.	Morale of teachers	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032502
49.	Students' attitudes toward academic achievement	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032503
50.	Teachers' attitudes toward academic achievement	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032504
51.	Parental support for student achievement	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032505
52.	Regard for school property	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032506
53.	Relations between students and teachers	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032507

Questions 54 – 58. To what extent has each of the following served as an impetus to change in the curriculum or instructional practices within your school during the past five years? (Answer only if you have been at the school or district for at least two years.) Fill in **one** oval on each line.

		To a great extent	To some extent	Not at all	
54.	District or school testing programs	Ⓐ	Ⓑ	Ⓒ	C032602
55.	State testing mandates	Ⓐ	Ⓑ	Ⓒ	C032604
56.	Public reporting of school or district performance data	Ⓐ	Ⓑ	Ⓒ	C032606
57.	Budget changes	Ⓐ	Ⓑ	Ⓒ	C032609
58.	Changes in student body or in student assignment policies	Ⓐ	Ⓑ	Ⓒ	C032610

59. Are minimum requirements for time spent on homework in effect in your school this year? WP000050

Ⓐ Yes

Ⓑ No

Questions 60 – 66. To what degree is each of the following a problem in your school? Fill in one oval on each line.

	Serious	Moderate	Minor	Not a problem	
60. Student tardiness	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032401
61. Student absenteeism	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032402
62. Student cutting of classes	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032403
63. Physical conflicts among students	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032404
64. Racial or cultural conflicts	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032407
65. Student health problems	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032408
66. Teacher absenteeism	Ⓐ	Ⓑ	Ⓒ	Ⓓ	C032406

67. About what percentage of your students is absent on an average day? (Include excused and unexcused absences in calculating this rate.) C033600

Ⓐ 0-2%

Ⓑ 3-5%

Ⓒ 6-10%

Ⓓ More than 10%

68. About what percentage of your teachers is absent on an average day? (Include excused and unexcused absences in calculating this rate.) LC000488

Ⓐ 0-2%

Ⓑ 3-5%

Ⓒ 6-10%

Ⓓ More than 10%

69. About what percentage of students who are enrolled at the beginning of the school year is still enrolled at the end of the school year? (Exclude students who transfer into the school during the school year in figuring this rate.) C033700

Ⓐ 98-100%

Ⓑ 95-97%

Ⓒ 90-94%

Ⓓ Less than 90%

70. Of the full-time teachers who started the 1992-93 school year in your school, about what percentage left before the end of the school year? C033903

Ⓐ 0%

Ⓑ 1-2%

Ⓒ 3-5%

Ⓓ 6-10%

Ⓔ More than 10%

71. Of the students enrolled in the 12th grade in 1992-93, about what percentage was retained in the 12th grade in 1993-94? LC000517

Ⓐ 0%

Ⓑ 1-2%

Ⓒ 3-5%

Ⓓ 6-10%

Ⓔ More than 10%

Questions 72 – 76. How many of the following types of specialists or aides work in your school? Fill in one oval on each line.

	None	Less than one full-time equivalent	One-Two	Three-Four	Five or more	
72. Counselors	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	C034007
73. Psychologists	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	C034008
74. Social workers	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	C034009
75. Full-time librarian	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	LC000494
76. Media specialist	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	LC000495

77. Which of the following best describes the primary way in which your library is staffed? LC000502

Ⓐ No library in school

Ⓑ Library in school, no staff or only volunteer staff available

Ⓒ Part-time staff

Ⓓ Full-time staff

Questions 78 – 79. Of students in last year's graduating class, approximately what percentage has gone on to attend each of the following? Fill in **one** oval on each line.

	0-10%	11-25%	26-50%	51-75%	76-90%	91-100%	
78. Two-year colleges or universities	(A)	(B)	(C)	(D)	(E)	(F)	C036001
79. Four-year colleges or universities	(A)	(B)	(C)	(D)	(E)	(F)	C036002

80. What is/are the title(s) of the person or persons who filled out this questionnaire? Fill in all ovals that apply. HE000325

(A) Principal

(B) Headmaster/Headmistress

(C) Head teacher

(D) Vice Principal, Assistant Principal

(E) Counselor

(F) Curriculum Coordinator, Department Head

(G) Teacher

(H) Secretary

(I) Other C034100

81. Does your school receive Chapter I funding? WP000069

(A) Yes

(B) No

82. What percentage of your students are Chapter I eligible? WP000070

(A) 10% or below

(B) 11-25%

(C) 26-75%

(D) 76-99%

(E) 100%

APPENDIX C

SCHOOL INFORMATION FORM

NAEP SCHOOL ID: _____

SUPERVISOR: _____

SCHOOL INFORMATION FORM
1994 HIGH SCHOOL TRANSCRIPT STUDY

A. SCHOOL INFORMATION

SCHOOL NAME: _____

CITY, STATE: _____

PRINCIPAL: _____ TELEPHONE: (____) _____

1. WHO WILL BE THE SCHOOL COORDINATOR FOR THE HSTS? Name: _____

CIRCLE EITHER 1 OR 2

SAME PERSON AS 1994 NAEP 1
NEW PERSON 2

RECORD NAME AND PHONE NUMBER:

NAME: _____
TELEPHONE: _____

DOES THE COORDINATOR WORK IN THE SUMMER?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, AVAILABLE WHEN? DATES: _____

HOURS: _____

2. SCHOOL HOURS: _____

3. SUMMER OFFICE HOURS:

DATES: _____

HOURS: _____

4. LAST DAY OF SCHOOL: _____
Date

5. 1994 GRADUATION DATE: _____
Date

6a. WHEN WILL THE TRANSCRIPTS FOR THE 1994 GRADUATES BE AVAILABLE? _____
Date

6b. WHEN WOULD BE THE MOST CONVENIENT TIME FOR SOMEONE TO RETURN TO GET COPIES OF TRANSCRIPTS? _____
Date

7. 1994-95 SCHOOL YEAR BEGINS: _____
Date

IF DISTRICT/SCHOOL REFUSES TO PARTICIPATE, EXPLAIN:

8. WHERE AND WITH WHOM WILL THE SCHOOL'S COPY OF THE 1994 NAEP ADMINISTRATION SCHEDULE(S) BE KEPT?

9. EXPLAIN TO COORDINATOR THE SYSTEM FOR INSERTING DISCLOSURE NOTICES IN STUDENT FILES AND OBTAINING TRANSCRIPTS AFTER GRADUATION. BE SURE TO DISCUSS THAT NO STUDENT TIME IS INVOLVED, CONFIDENTIALITY IS MAINTAINED, AND TRANSCRIPT REIMBURSEMENT IS PROVIDED.

COMMENTS ABOUT OBTAINING TRANSCRIPTS:

_____

_____

_____

_____

_____

10. WHO FILLED OUT THE IEP/LEP QUESTIONNAIRE?

CHECK ALL THAT ARE APPLICABLE:

FOR STUDENTS WITH IEP:

___ SPECIAL EDUCATION TEACHER/COORDINATOR
___ REGULAR EDUCATION TEACHER
___ GUIDANCE COUNSELOR
___ OTHER (SPECIFY) _____

FOR STUDENTS WITH LEP:

___ ESL TEACHER/COORDINATOR
___ REGULAR CLASSROOM TEACHER
___ GUIDANCE COUNSELOR
___ FOREIGN STUDENT COORDINATOR
___ OTHER (SPECIFY) _____

11. EXPLAIN TO COORDINATOR THAT YOU WANT COURSE CATALOGS FOR YEARS 90-91, 91-92, 92-93, AND 93-94. CATALOGS SHOULD CONTAIN ALL COURSES, INCLUDING VOCATIONAL HONORS, REMEDIAL, SPECIAL ED., AND OFF-CAMPUS. EXPLAIN THE TYPES OF CATALOGS NEEDED IN ORDER OF PREFERENCE AS FOLLOWS:

- School-level catalogs that provide course names and descriptions;
- District-level catalogs that provide course names and descriptions with the course offerings for this particular school clearly indicated;
- A course list by department that includes general descriptions of course offerings by department;
- A course list by department that includes general descriptions of course offerings by department;
- Course lists without descriptions;
- District-level catalogs without school-level indication.

ARE CATALOGS AVAILABLE NOW?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF NO, WHEN WILL THEY BE AVAILABLE?_____
pick-up date

COMMENTS ABOUT OBTAINING COURSE CATALOGS:

_____

_____

_____

12. EXPLAIN THAT YOU WOULD LIKE TO HAVE A SAMPLE OF THREE TRANSCRIPTS FOR STUDENTS WHO HAVE ALREADY GRADUATED (WITHOUT NAMES OR IDENTIFYING INFORMATION). THE SAMPLE TRANSCRIPTS SHOULD REFLECT REGULAR COURSES, HONORS COURSES, AND SPECIAL EDUCATION COURSES.

13. IF COORDINATOR MENTIONS NEED FOR PARENTAL CONSENT, SHOW FERPA, NCES LETTERS AND, IF NECESSARY, PARENTAL CONSENT LETTERS. RECORD COORDINATOR'S REACTIONS.

14. ESTABLISH APPOINTMENT TO GET CATALOGS AND TRANSCRIPTS, AS APPROPRIATE.

B. OBTAINING COURSE CATALOGS

1. CHECK WHICH TYPE(S) OF CATALOGS OBTAINED

___ School-level catalogs that provide course names and descriptions

___ District-level catalogs that provide course names and descriptions with the course offerings for this particular school clearly indicated

___ A course list by department that includes general descriptions of course offerings by department

___ A course list by department that includes general descriptions of course offerings by department

___ Course lists without descriptions

___ District-level catalogs without school-level indication

ON THE LINES BELOW, RECORD THE TITLE OF EACH CATALOG YOU RECEIVE. RECORD THE SCHOOL ID AND CATALOG # ON THE COVER OF THE DOCUMENT.

CATALOG #	TITLE
1	_____
2	_____
3	_____
4	_____
5	_____
6	_____
7	_____
8	_____
9	_____
10	_____

2. COMPLETE THE HSTS COURSE CATALOG CHECKLIST.

3. THE CATALOGS YOU OBTAIN SHOULD COVER ALL COURSES AVAILABLE TO THE CLASS OF 1994 DURING ALL THEIR YEARS AT THIS SCHOOL (INCLUDING 9TH GRADE COURSES IF TAKEN AT A JUNIOR HIGH/MIDDLE SCHOOL).

A. DO THEY INCLUDE VOCATIONAL COURSES?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, HOW ARE THEY IDENTIFIED IN THE CATALOG(S)?

_____

_____

_____

B. DO THEY INCLUDE REMEDIAL COURSES?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, HOW ARE THEY IDENTIFIED IN THE CATALOG(S)? _____

_____

_____

_____

C. DO THEY INCLUDE "HONORS" COURSES?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, HOW ARE THEY IDENTIFIED IN THE CATALOG(S)? _____

_____

_____

_____

D. DO THEY INCLUDE SPECIAL ED. COURSES?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, ARE DIFFERENT LEVELS OF SPECIAL ED. IDENTIFIED (I.E., RESOURCE AND SELF-CONTAINED CLASSES)?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, HOW ARE THEY IDENTIFIED? _____

_____

E. DO THEY INCLUDE OFF-CAMPUS COURSES?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, HOW ARE THEY IDENTIFIED IN THE CATALOG(S)?

_____

_____

_____

F. DO THEY INCLUDE ESL OR BILINGUAL COURSES? (COURSES TAUGHT IN A LANGUAGE OTHER THAN ENGLISH)

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, HOW ARE THEY IDENTIFIED IN THE CATALOG(S)? _____

_____

_____

_____

4. |__| COURSE CATALOG CHECKLIST COMPLETED.

5. IF WESTAT STAFF HAVE QUESTIONS ABOUT THE COURSE CATALOGS, WHO IS THE BEST PERSON TO CONTACT?

|__| SCHOOL COORDINATOR

OTHERS (NAME)	TITLE	PHONE
_____	_____	_____
_____	_____	_____

C. OBTAINING OTHER SCHOOL INFORMATION

1. FOR 1993-94, HOW MANY CREDITS DOES A STUDENT IN THIS SCHOOL EARN FOR A COURSE TAKEN FOR A SINGLE CLASS PERIOD, THAT LASTS FOR THE WHOLE SCHOOL YEAR?

_____
OF CREDITS

1a. HAS THIS CHANGED DURING THE LAST FOUR SCHOOL YEARS?

CIRCLE EITHER 1 OR 2

YES 1 (Q1b)
NO 2

1b. IF YES, HOW MANY CREDITS WERE GIVEN FOR A YEAR-LONG COURSE IN EACH OF THOSE YEARS?

1990-91 _____
CREDITS

1991-92 _____
CREDITS

1992-93 _____
CREDITS

2. HOW MANY CLASS PERIODS DOES A TYPICAL 12TH GRADER HAVE PER DAY, NOT COUNTING LUNCH?

_____
OF CLASS PERIODS

3. WHAT IS THE MAXIMUM NUMBER OF CLASS PERIODS A STUDENT IN THIS SCHOOL MAY TAKE EACH DAY?

_____
MAX. # OF CLASS PERIODS

4. WHAT IS THE MINIMUM NUMBER OF CLASS PERIODS ALL STUDENTS IN THIS SCHOOL MAY TAKE EACH DAY?

_____
MIN. # OF CLASS PERIODS

5. IS THE MINIMUM NUMBER OF COURSES DIFFERENT FOR SENIORS?

YES 1
NO 2

6. HOW LONG DOES THE TYPICAL CLASS PERIOD LAST?

_____
MINUTES

7. ARE CREDITS FOR HONORS/AP COURSES DEFINED THE SAME AS ABOVE?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF NO, DESCRIBE ANY DIFFERENCES _____

_____

_____

8. ARE CREDITS FOR SPECIAL EDUCATION STUDENTS DEFINED THE SAME AS ABOVE?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF NO, EXPLAIN THE DIFFERENCE: _____

_____

_____

9. DOES THIS SCHOOL INCLUDE 9TH GRADE?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

9a. IF YES, DO MOST STUDENTS ATTEND THIS SCHOOL FOR FOUR YEARS, INCLUDING 9TH GRADE?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

9b. IF NO, WHERE DO MOST STUDENTS ATTEND 9TH GRADE?

___ A SINGLE FEEDER JUNIOR HIGH/MIDDLE SCHOOL
___ SEVERAL JUNIOR HIGH/MIDDLE SCHOOLS IN THE DISTRICT
___ OTHER SCHOOLS NOT IN THIS DISTRICT OR AFFILIATED WITH THIS SCHOOL

10. WHAT TYPES OF DIPLOMAS ARE OFFERED?

___ Standard
___ Regents (NY State only)
___ Honors
___ Certificate of Merit
___ Vocational
___ Special Education
___ Certificate of Attendance
___ International Baccalaureate
___ Other (PLEASE DESCRIBE)

11. WE NEED TO KNOW THE GRADUATION REQUIREMENTS FOR ALL HIGH SCHOOL DIPLOMA PROGRAMS OFFERED AT THIS SCHOOL. IF THIS IS DOCUMENTED IN THE COURSE CATALOG(S), CHECK THE BOX BELOW AND INDICATE WHERE. PLACE A PAPER CLIP ON CATALOG PAGES WHERE GRADUATION REQUIREMENTS ARE DESCRIBED. OTHERWISE, CONTINUE WITH Q12.

|__| GRADUATION REQUIREMENTS RECORDED ON PAGE(S): _____

(SKIP TO Q13)

12. WHAT ARE THE GRADUATION REQUIREMENTS FOR (DIPLOMA TYPE) IN THE FOLLOWING SUBJECT AREAS? (CHECK BOX IF NOT OFFERED.)

SUBJECT AREAS	STANDARD Credits	HONORS N/A ☐ Credits	VOCATIONAL N/A ☐ Credits	OTHER (_____) N/A ☐ Credits
a. English/Language Arts	_____	_____	_____	_____
b. Mathematics	_____	_____	_____	_____
c. Computer Science	_____	_____	_____	_____
d. Social Studies/History	_____	_____	_____	_____
e. Science	_____	_____	_____	_____
f. Foreign Language	_____	_____	_____	_____
g. Physical Education/Health	_____	_____	_____	_____
h. OTHER (_____)	_____	_____	_____	_____
i. OTHER (_____)	_____	_____	_____	_____
TOTAL CREDITS REQUIRED FOR GRADUATION*	_____	_____	_____	_____

*This number may be larger or smaller than the credits specified for A-I above because of electives and/or overlapping areas.

13. ARE THERE ANY COURSES REQUIRED FOR GRADUATION THAT DO NOT RECEIVE CREDITS? IF YES, SPECIFY _____

_____

14. DO THESE GRADUATION REQUIREMENTS ASSUME FOUR YEARS OF HIGH SCHOOL?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF NO, EXPLAIN: _____

_____

_____

15. ARE THERE GRADE REQUIREMENTS FOR GRADUATION?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, EXPLAIN: _____

_____

_____

16. ARE THERE STATE OR DISTRICT COMPETENCY TESTS THAT ARE REQUIRED FOR GRADUATION?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF YES, EXPLAIN: _____

_____

_____

17. IF WESTAT STAFF HAVE QUESTIONS ABOUT CREDITS, GRADUATION REQUIREMENTS, ETC., WHO IS THE BEST PERSON TO CONTACT?

|__| SCHOOL COORDINATOR

OTHERS (NAME)	TITLE	PHONE
_____	_____	_____
_____	_____	_____

D. REVIEWING THE TRANSCRIPTS

COMPLETE THIS SECTION WHILE YOU ARE AT THE SCHOOL AND AFTER YOU HAVE RECEIVED COPIES OF THE SAMPLE TRANSCRIPTS.

SAMPLE TRANSCRIPTS OBTAINED INCLUDE:

CHECK ALL THAT APPLY:

___ Regular courses
___ Honors courses
___ Special education courses

1. COMPLETE TRANSCRIPT FORMAT CHECKLIST

2. IS THE TYPICAL "A, B, C" GRADING SYSTEM USED?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF NO, EXPLAIN THE GRADING SYSTEM: _____

_____

_____

3. IS THE GRADING SYSTEM THE SAME FOR ALL STUDENTS (I.E., SPECIAL EDUCATION, HONORS, ETC.?)

CIRCLE EITHER 1 OR 2

YES 1
NO 2

IF NO, EXPLAIN: _____

_____

_____

4. DO COURSE TITLES OR COURSE NUMBERS ON THE TRANSCRIPTS MATCH THOSE IN THE COURSE CATALOG?

CIRCLE EITHER 1 OR 2

YES 1
NO 2

COMMENTS:_____

_____

_____

5. IF THERE ARE ABBREVIATIONS OR SYMBOLS ON THE TRANSCRIPTS WHICH ARE NOT SELF-EVIDENT, FIND OUT WHAT THEY STAND FOR AND RECORD ON THE TRANSCRIPT FORMAT CHECKLIST.

6. FINAL SAMPLE TRANSCRIPT CHECKLIST:

A. |__| ALL CHECKED FOR LEGIBILITY AND COMPLETENESS

B. |__| NAMES AND IDENTIFIERS HAVE BEEN REMOVED FROM EACH

C. |__| TRANSCRIPT FORMAT CHECKLIST COMPLETED

7. IF WESTAT STAFF HAVE QUESTIONS ABOUT THE TRANSCRIPTS, WHO IS THE BEST PERSON TO CONTACT?

|__| SCHOOL COORDINATOR

OTHERS (NAME)	TITLE	PHONE
_____	_____	_____
_____	_____	_____

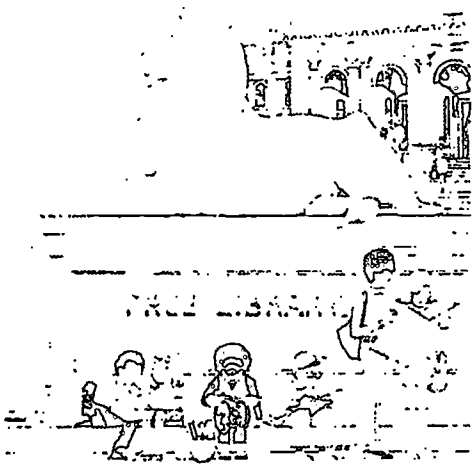
APPENDIX D

1993-94 IEP/LEP STUDENT QUESTIONNAIRE

Teacher of:

93-94 IEP/LEP Student Questionnaire

Q-094



GRADE

1 2 3 4 5 6 7 8 9 10 11 12

BIRTH DATE

Month: JAN (1) FEB (2) MAR (3) APR (4) MAY (5) JUN (6) JUL (7) AUG (8) SEP (9) OCT (10) NOV (11) DEC (12)

Year: (6) (7) (8) (9)

SEX

M F

THE NATION'S REPORT CARD

SCHOOL #

R

IEP

LEP

Chap 1

DO NOT USE

STUDENT SELECTED FOR:

STATE

READING

NATIONAL

READING

HISTORY/GEOGRAPHY

A project of the Office of Educational Research and Improvement.
This report is authorized by law (20 USC 1221 e-1(b)). While you are not required to respond, your cooperation is needed to make the results of the survey comprehensive, accurate, and timely.

OMB NO. 1850-0628 • Approval Expires 6/94
Mark Reflex® by NCS EP-153906-001:3 Printed in U.S.A.

IEP/LEP Student Questionnaire

During the 1993-94 school year, a sample of students across the country, including some students from your school, will be given a series of questions as part of the National Assessment of Educational Progress (NAEP). The current assessment focuses on achievement in reading, history, and geography. As part of the assessment, NAEP will investigate the relationship between students' achievement and various school, teacher, and home factors that may influence this achievement. In order to obtain a complete picture of how all children are doing, it is necessary to collect information on those students who have been identified as having an Individualized Education Program (IEP) or Limited English Proficiency (LEP) and are either assessed or NOT. We are asking you to complete this questionnaire about one of those students.

We realize you are very busy; however, we urge you to complete this questionnaire as carefully as possible. The information you provide will be kept confidential.

NAEP is authorized under Public Law 100-297. While your participation is voluntary, your responses to these questions are needed to make this survey accurate and complete.

Please answer directly on the questionnaire by filling in the appropriate oval or by writing your response in the space provided. When you are finished, please return the questionnaire to your school's NAEP coordinator.

Thank you very much for your help.

1. Why is this student classified as IEP/LEP?

- (A) A disability (physical or mental disability)
  (PLEASE FILL IN SECTIONS A AND B)
- (B) Limited English proficiency
  (PLEASE FILL IN SECTIONS A AND C)
- (C) Both a disability and limited English proficiency
  (PLEASE FILL IN SECTIONS A, B, AND C)
- (D) Nonreader but does not have a disability or limited English proficiency
  (PLEASE DO NOT FILL IN THE REST OF THE QUESTIONNAIRE)
- (E) Other reason (specify)_____
  (PLEASE DO NOT FILL IN THE REST OF THE QUESTIONNAIRE)

Please continue on next page ▷

Section A: Functional Grade Level and Mainstreaming

(Complete this section if this student has a disability and/or has limited English proficiency.)

2. What functional grade level has this student achieved in reading English?

- (A) No grade level determined
- (B) Lower than kindergarten
- (C) Kindergarten
- (D) Grade 1
- (E) Grade 2
- (F) Grade 3
- (G) Grade 4
- (H) Grade 5
- (I) Grade 6
- (J) Grade 7
- (K) Grade 8
- (L) Grade 9
- (M) Grade 10
- (N) Grade 11
- (O) Grade 12

X004301

3. What functional grade level has this student achieved in mathematics?

- (A) No grade level determined
- (B) Lower than kindergarten
- (C) Kindergarten
- (D) Grade 1
- (E) Grade 2
- (F) Grade 3
- (G) Grade 4
- (H) Grade 5
- (I) Grade 6
- (J) Grade 7
- (K) Grade 8
- (L) Grade 9
- (M) Grade 10
- (N) Grade 11
- (O) Grade 12

X004401

4. What percentage of the school day does this student spend in a regular class (i.e., mainstream) setting?

- (A) 0%
- (B) 10%
- (C) 20%
- (D) 30%
- (E) 40%
- (F) 50%
- (G) 60%
- (H) 70%
- (I) 80%
- (J) 90%
- (K) 100%

X004501

Section B: Students With a Disability

(Complete this section if this student has a disability)

5. Which of the following best describes this student's disability?

- (A) Multidisabled
- (B) Mentally retarded
- (C) Hard of hearing
- (D) Deaf
- (E) Speech-impaired
- (F) Visually handicapped/blind
- (G) Deaf/blind
- (H) Emotionally disturbed
- (I) Orthopedically impaired
- (J) Learning disabled
- (K) Other (specify) _____

X005601

6. How would you describe this student's condition?

- (A) Profound
- (B) Severe
- (C) Moderate
- (D) Mild

X004601

7. What percentage of the school day is this student served by a special education program?

- (A) 0%
- (B) 10%
- (C) 20%
- (D) 30%
- (E) 40%
- (F) 50%
- (G) 60%
- (H) 70%
- (I) 80%
- (J) 90%
- (K) 100%

X004701

Please continue on next page. ▷

Questions 8-14. Is this student currently receiving instruction in any of the following areas as part of a special education program? Fill in one oval on each line.

	Yes	No
8. Language development	A	B
9. Reading	A	B
10. Mathematics	A	B
11. Speech (e.g., articulation, voice, speech flow)	A	B
12. Self-control and deportment	A	B
13. Personal care and basic life skills	A	B
14. Vocational education	A	B

X004600

Section C: Students With Limited English Proficiency

(Complete this section if this student has limited English proficiency.)

15. What is this student's non-English language?

- A Spanish
- B Another language (specify) _____

X004901

16. What percent of the students in this school speak this student's non-English language?

- A None
- B 10% or less
- C 11-20%
- D 21-30%
- E 31-40%
- F 41-50%
- G 51-60%
- H More than 60%

X005001

17. Last year did this student live in a territory or country where English is *not* the dominant language?

- A Yes
- B No
- C I don't know

X005101

18. What percentage of the school day is this student served by a special language program?

A 0%	E 40%	I 80%
B 10%	F 50%	J 90%
C 20%	G 60%	K 100%
D 30%	H 70%	

X005201

Questions 19-21. Is this student currently receiving any of the following types of instruction as a part of a special language program? Fill in one oval on each line.

	Yes	No
19. English language course designed for speakers of another language	A	B
20. A course in reading and writing in the student's native language	A	B
21. One or more content courses (eg., mathematics, science, social studies) taught in the student's native language	A	B

X005300

22. Counting this year, how many years has this student been in a special language program?

- A Student is not in a special language program.
- B 1 year
- C 2 years
- D More than 2 years
- E I don't know

X005401

Questions 23-26. How would you characterize this student's proficiency in English? Fill in one oval on each line.

	Excellent	Good	Fair	Poor	No proficiency	I don't know
23. Speaking	A	B	C	D	E	F
24. Understanding	A	B	C	D	E	F
25. Reading	A	B	C	D	E	F
26. Writing	A	B	C	D	E	F

X005500

Thank you for your cooperation.

APPENDIX E

1994 ADDITIONS TO THE CLASSIFICATION OF SECONDARY SCHOOL COURSES

APPENDIX E

1994 ADDITIONS TO THE CLASSIFICATION OF SECONDARY SCHOOL COURSES

16.1200	Indo-European Languages, Other
16.1300	Non-English Languages for Native Speakers, Other
21.0127	Intro to Technology Technology Education general course on technology, including computers, computer-related and computer-controlled technology
27.0425	Geometry, Part 1 geometry 1 taught over 2 years; 1st year full credit
27.0426	Geometry, Part 2 geometry 1 taught over 2 years; 2nd year full credit
27.0427	Unified Math 1, Part 1 Unified math taught over 2 years; 1st year full credit
27.0428	Unified Math 1, Part 2 unified math taught over 2 years; 2nd year full credit
32.0231	Individualized Academic Program dropout prevention, college preparation, tutorial assistance, e.g. project AVID
35.0141	Dropout Prevention communities/cities in schools
42.0114	AP Psychology
45.0613	AP Economics
48.0271	Desktop Publishing

ISBN 0-16-049170-3

193

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Standard Mail (B)

NATIONAL CENTER FOR EDUCATION STATISTICS

ERRATA

The 1994 High School Transcript Study Technical Report

The enclosed pages 6–15, 6–40, 6–41, 6–42, 6–43, 6–44, 6–50, and 6–52 are replacement pages for the corresponding pages in ***The 1994 High School Transcript Study Technical Report*** (NCES 97–262). The estimates presented in these tables and surrounding text reflect a change in the post-stratification weighting of estimates used to report the transcript results.

October 14, 1998

U.S. Department of Education
Office of Educational Research and Improvement **NCES 97–262e**

$$w_{ij}(r) = \begin{cases} w_{ij} & i \notin S(r) \\ 2w_{ij} & i \in S(r,1) \\ 0 & i \in S(r,2) \end{cases} \qquad r=1,\ldots,25,\ r=28,\ldots,36.$$

In the special case of the "triplet" of PSUs corresponding to r=26 one of the PSUs was randomly assigned to random half sample group 1, S(26,1), and one to random half sample group 2, S(26,2). The remaining PSU is designated as S(26,3). The replicate weights assigned for r=26 are then as follows:

$$w_{ij}(26) = \begin{cases} w_{ij} & i \notin S(26) \\ 1.5w_{ij} & i \in S(26,1) \\ 0 & i \in S(26,2) \\ 1.5w_{ij} & i \in S(26,3) \end{cases}$$

For r=37,....62, the replicate weights correspond to certainty PSUs. The replicate groups for these replicate weights correspond to sets of schools rather than to PSUs, as schools are the first stage sampling units for certainty PSUs (see Section 6.2). Write *S(r)* as the set of schools corresponding to replicate weight r.[15] The replicate groups are generated by randomly assigning one of the half sample groups of schools to random half sample group 1 for pairs, and randomly assigning two of the three groups of schools to random groups 1 and 2 for triplets.[16] (See also Section 6.2). These random half sample groups will be indicated as S(r,1) and S(r,2), with an S(r,3) also for the triplets. After this random selection has taken place the replicate school base weight for the j-th school, i-th PSU, r-th replicate weight, is computed as follows for the pairs:

$$w_{ij}(r) = \begin{cases} w_{ij} & ij \notin S(r) \\ 2w_{ij} & ij \in S(r,1) \\ 0 & ij \in S(r,2) \end{cases} \qquad r=37,\ldots,62,\ r\neq 42,52,54,57,58,59,60,62.$$

For the replicate weights corresponding to triplets (PSUs with three HSTS sample schools), the computation of base weights is more elaborate. Each of these PSUs is assigned a companion PSU as follows: 42 and 52, 54 and 62, 57 and 58, and 59 and 60. The assignment of

[15] For all but two of the certainty PSUs, this set corresponds to all schools in the PSU. For the Los Angeles and New York PSUs, this set for each replicate weight corresponds to a third of the schools in the PSU.

[16] As discussed in Section 6.2, these "triplet" certainty PSUs had three HSTS schools. These PSUs correspond to 8 of the replicate weights.

to average performance over a conceptually infinite number of replications of the sampling, it is unlikely that any given estimate, based on the achieved sample, will exactly equal the population value. Furthermore, the respondent weights have been adjusted for nonresponse and a few extreme weights have been reduced in size.

To reduce the mean square error of estimates using the sampling weights, these weights will be further adjusted so that estimated population totals for a specified subgroup population, based on the sum of student weights for a specified type, will be the same as presumably better estimates based on composites of estimates from the Current Population Survey. This adjustment, called poststratification, is intended especially to reduce the mean squared error of estimates relating to student populations that span several subgroups of the population. The poststratification classes are defined in terms of race/ethnicity and NAEP region.

For the HSTS weights, the post-stratification adjustment factor ($STPSAF_g$) for the g^{th} post-stratification adjustment cell will be:

$$STPSAF_g = \frac{C_g}{\sum_{ijk \in E(g)} w_{ijk} \, SCNRAF_{ij} \, I[ij \in SR(\alpha)] \, TRIM(ij)} \qquad \text{(Equation 6.6.16)}$$

The quantity C_g is the 12th grade enrollment control total of students whose 18th birthday was on or after October 1, 1993 for the g^{th} poststratification class. E(g) is the collection of all students in the g^{th} poststratification class who were enrolled in 12th grade (including those who did not graduate in 1994) and whose 18th birthday was on or after October 1, 1993. The quantity w_{ijk} is the full sample student base weight for the k^{th} student in the j^{th} school in the i^{th} PSU, that was discussed in Section 6.4.1. The final three factors comprise the school nonresponse adjustment for the HSTS weights, discussed in Section 6.5.4., and the trimming factor for the school, discussed in Section 6.6.2.

Table 6 14 presents the poststratification cells with the CPS control totals for each cell. Control totals are given in thousands. For a discussion of the definition of regions as used in NAEP, see Section 2.2.

Table 6-14. Student poststratification cells and control totals

Poststratification Cell	Race/Ethnicity	Region	CPS control total (000)
1	Blacks, nonHispanic	All	235.3
2	Hispanics	All	159.2
3	Other race, nonHispanic	All	102.6
4	Whites, nonHispanic	Northeast	347.0
5	Whites, nonHispanic	Southeast	342.8
6	Whites, nonHispanic	Central	494.7
7	Whites, nonHispanic	West	414.5

Table 6-15 presents the aggregated weights within each poststratification cell (the denominator of Equation 6.6.16), the control total C_g, and the poststratification factor $STPSAF_g$ for the poststratification cell.

Table 6-15. HSTS poststratification factors

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification factor
1	204.6	235.3	1.150
2	177.0	159.2	0.899
3	121.6	102.6	0.843
4	332.0	347.0	1.045
5	322.2	342.8	1.064
6	408.6	494.7	1.211
7	334.7	414.5	1.238

In Table 6-15 and the remaining tables in Section 6.6, the poststratification factor as given is the unrounded control total divided by the unrounded aggregated weight. The control totals and aggregated weights given in the tables are the corresponding total rounded to one digit after the decimal point. The poststratification factor as given may not equal the ratio of the two rounded summands as given in all cases.

6.6.5 Poststratification for the Linked Weights: Assessment Weights

The poststratification procedure is similar to the corresponding procedure for the HSTS weights as described in Section 6.6.4, in that the same poststratification categories and control totals are used. In this case, however, separate adjustments are made for each of the three assessments, and for the excluded students.

For the three assessments, each assessment sample must represent the full population. The control totals however are not separable into students eligible for an assessment, and excluded students. Because of this nonseparability, the excluded students from the sample must be included with the assessment group when computing the poststratification adjustment. For each assessment a the poststratification factor corresponding to poststratification class g is as follows:

$$STPSFL_{ag} = \frac{C_g}{\left[\sum_{ijk \in E(g),\ ijk\ assessed} w^a_{ijk}\ SCNRFL_{a\alpha}\ I[ij \in SRL_a(\alpha)]\ STNNRF_{a\gamma}\ I[ijk \in STR_a(\gamma)]\ TRIM_a(ij) + \sum_{ijk \in E(g),\ ijk\ excluded} w^e_{ijk}\ SCNRFL_{a\alpha}\ I[ij \in SRL(\alpha)]\ TRIM_a(ij) \right]}$$

(Equation 6.6.17)

The quantity C_g in the numerator of Equation 6.6.17 represents the 12th-grade enrollment control total of students whose 18th birthday was on or after October 1, 1993 for the g^{th} poststratification class. E(g) is the collection of all students in the g^{th} poststratification class who were enrolled in 12th-grade (including those who did not graduate in 1994) and whose 18th birthday was on or after October 1, 1993. The quantities w^a_{ijk} and w^e_{ijk} are the student base weights for assessed and excluded students respectively, discussed earlier in Section 6.4.3.

There are school nonresponse adjustment factors in both the assessed and excluded student summations, discussed in Section 6.5.5, and student nonresponse adjustment factors for the assessed students only, discussed in Section 6.6.1. The final factors in each term of each summation are trimming factors for the weights, discussed in Section 6.6.3.

Tables 6-16, 6-17, and 6-18 present the aggregated weights (the denominator of Equation 6.6.17), the control totals C_g, and the poststratification factors $STPSFL_{ag}$ for each poststratification cell for the reading assessment, the history assessment, and the geography assessment, respectively.

Table 6-16. Poststratification factors for the reading assessment weights

Poststratification Cell	Aggregated weight (000)	Control total (000)	Poststratification factor (000)
1	205.0	235.3	1.148
2	180.4	159.2	0.882
3	133.6	102.6	0.768
4	349.6	347.0	0.992
5	306.4	342.8	1.119
6	412.7	494.7	1.199
7	336.7	414.5	1.231

Table 6-17. Poststratification factors for the history assessment weights

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification factor (000)
1	209.2	235.3	1.125
2	176.1	159.2	0.904
3	121.2	102.6	0.846
4	349.8	347.0	0.992
5	327.7	342.8	1.046
6	410.7	494.7	1.204
7	311.8	414.5	1.329

Table 6-18. Poststratification factors for the geography assessment weights

Poststratification cell	Aggregated weight (000)	Control total (000)	Poststratification Factor (000)
1	217.5	235.3	1.082
2	180.5	159.2	0.882
3	144.9	102.6	0.708
4	338.6	347.0	1.025
5	326.1	342.8	1.051
6	397.0	494.7	1.246
7	329.0	414.5	1.260

6.6.6 Special Poststratification Adjustments for the Final Excluded Student Weights

The poststratification adjustment for the excluded students needs to include all students, since control totals do not exist for excluded students alone. In this case, all students from all of the assessments are included, along with the excluded students, when computing the adjustments. The weights used for these students are not the weights adjusted for selection into an assessment. Rather, they are the original weights reflecting selection into the HSTS sample: the original HSTS base weights adjusted for school nonresponse (using the excluded student linked weight adjustments).

The poststratification adjustment for excluded students is shown as follows:

$$STPSFL_{eg} = \frac{C_g}{\left[\sum_{ijk \in E(g),\ ijk\ assessed} w_{ijk} SCNRFL_{\alpha i}\, I[ij \in SRL(\alpha)]\, STNNRF_{\gamma}\, I[ijk \in STR(\gamma)]\, TRIM_S(ij) + \sum_{ijk \in E(g),\ ijk\ excluded} w_{ijk} SCNRFL_{\alpha i}\, I[ij \in SRL(\alpha)]\, TRIM_S(ij) \right]} \quad \text{(Equation 6.6.18)}$$

The school nonresponse adjustment factors were discussed in Section 6.5.5, student nonresponse adjustment factors in Section 6.6.1, and trimming factors in Section 6.6.3.

Table 6-19 presents the aggregated weights (the denominator of Equation 6.6.18), the control totals C_g, and the poststratification factors $STPSFL_{eg}$ for each poststratification cell.

Table 6-19. Poststratification factors for the excluded student weights

Poststratification Cell	Aggregated weight (000)	Control total (000)	Poststratification factor (000)
1	208 3	235.3	1.129
2	179.0	159.2	0.889
3	132.5	102.6	0.774
4	348.1	347.0	0.997
5	316.6	342.8	1.083
6	408.3	494.7	1.211
7	327.8	414.5	1.264

Table 6-23. Nonresponse adjustment factors for missing transcripts

Cell Number	$STWAF_h$	Reading assessment $STWFL_{ah}$	History assessment $STWFL_{ah}$	Geography assessment $STWFL_{ah}$	Excluded students $STWFL_{ah}$
1	1.031	1.026	1.027	1.011	1.175
2	1.010	1.012	1.005	1.005	1.175
3	1.025	1.028	1.018	1.029	1.175
4	1.057	1.012	1.022	1.028	1.496
5	1.024	1.019	1.014	1.015	1.496
6	1.138	1.050	1.034	1.012	1.217
7	1.019	1.034	1.012	1.011	1.217
8	1.129	1.076	1.061	1.051	1.217
9	1.017	1.020	1.004	1.012	1.217
10	1.042	1.043	1.023	1.016	1.204
11	2.060	1.000	1.000	1.000	1.204

6.7.3 Final Sampling Weights

Final HSTS sampling weights were assigned to students in the HSTS study for which a transcript was received. These sampling weights are computed as follows:

$$W_{ijk} = w_{ijk}\, SCNRAF_{\alpha}\, I[ij \in SR(\alpha)]\, TRIM(ij)\, STPSAF_{g}\, I[ijk \in E(g)] * STWAF_{h}\, I[ijk \in GR(h)] \quad \text{(Equation 6.7.4)}$$

The first factor is the student base weight, discussed in Section 6.4.1. The second and third factors comprise the school nonresponse adjustment, discussed in Section 6.5.4. The fourth factor is the school's trimming factor, discussed in Section 6.6.2. The fifth and sixth factors comprise the student poststratification factors, discussed in Section 6.6.4. Finally, the remaining two factors comprise the student missing transcript adjustment factor, discussed in Section 6.7.2.

Final linked sampling weights were assigned to all students in the HSTS study for which transcripts were received and who were assessed using one of the NAEP assessments. These weights are computed for each assessment a as follows:

$$W^{a}_{ijk} = w^{a}_{ijk}\, SCNRFL_{a\alpha}\, I[ij \in SRL_{a}(\alpha)]\, TRIM_{a}(ij)\, STNNRF_{a\gamma}\, I[ijk \in STR_{a}(\gamma)] * STPSFL_{ag}\, I[ijk \in E(g)]\, STWFL_{ah}\, I[ijk \in GR(h)] \quad \text{(Equation 6.7.5)}$$

Table 6-24. Distributions of the final HSTS and linked weights

Sample Distribution	HSTS weights	Reading assessment linked weights	History assessment linked weights	Geography assessment linked weights	Excluded student linked weights
Students with nonzero weights	25,335	9,258	5,070	4,143	533
Total (in thousands)	2,424	2,394	2,364	2,365	67.66
Minimum	1.60	26.77	17.52	11.24	20.59
25th percentile	53.60	150.01	269.59	325.94	80.78
Median	71.13	221.54	397.08	478.81	111.21
75th percentile	135.00	337.87	634.15	776.37	167.47
Maximum	659.21	2,510.42	1,632.87	2,340.4	277.21

6.7.4 Final Replicate Weights

The computation of final replicate school base weights is discussed in Section 6.4.5. It is only for this component that the replicate weights differ. The remaining weights and adjustments are computed as they were for the primary weights. The HSTS student base weights and student linked base weights are computed as follows:

$$w_{ijk}(r) = w_{ij}(r)w_{k|ij}$$
$$w^{a}_{ijk}(r) = w_{ijk}(r)w_{a|ij}w_{a|ijk}I[ijk \in U_a] \qquad \text{(Equation 6.7.7)}$$
$$w^{e}_{ijk}(r) = w_{ijk}(r)I[ijk \in U_e]$$

These quantities are defined in Sections 6.4.3 and 6.4.5. Note that all of these base weights are identical to the corresponding "main" (nonreplicate) base weights except for the factor $w_{ij}(r)/w_{ij}$.

In principle, the replicate weights should repeat the entire process of computing the final weights using the new replicate base weights. This replication will capture any components of variability introduced to the final weights by these processes. This was done for the HSTS and linked weights for most of these processes, except for the trimming step preceding poststratification, and the two CHAID analyses which selected school and missing transcript nonresponse cells.

The same trimming factors and CHAID categories were used for calculating the replicate weights as for the main weights. The components of variability introduced by these processes should be relatively small, so the complexity of replicating these processes led us to forgo replication of these