NCES
National Center for
Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2005-483

The High School Transcript Study

The 2000 High School Transcript Study User's Guide and Technical Report

The High School Transcript Study

The 2000 High School Transcript Study User's Guide and Technical Report

August 2005

U.S. Department of Education
Institute of Education Sciences
NCES 2005-483

Stephen Roey
Nancy Caldwell
Keith Rust
Lloyd Hicks
Janice Lee
Robert Perkins
Eyal Blumstein
Westat

Janis Brown
National Center for Education Statistics

Janis Brown
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Grover J. Whitehurst
Acting 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
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006

August 2005

The NCES World Wide Web Home Page is: http://nces.ed.gov
The NCES World Wide Web Electronic Catalog is: http://nces.ed.gov/pubsearch

Suggested Citation
Roey, S., Caldwell, N., Rust, K., Hicks, L., Lee, J., Perkins, R., Blumstein, E., and Brown, J. (2005). *The 2000 High School Transcript Study User's Guide and Technical Report* (NCES 2005–483). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

For ordering information on this report, write:

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

Call toll free 1–877–4ED–Pubs; or order online at *http://www.edpubs.org*

Content Contact:
Janis Brown
(202) 502–7482
janis.brown@ed.gov

ACKNOWLEDGMENTS

This report is the end product of the collaborative efforts of many individuals at the National Center for Education Statistics (NCES), Westat, and the American Institute for Research/Education Statistical Services Institute (ESSI) who contributed greatly to the High School Transcript Study (HSTS) project for the National Assessment of Educational Progress (NAEP) program. The combined talents, knowledge, and creativity of these individuals resulted in the successful completion of this report. Furthermore, the authors are indebted to all the schools and their staff that participated in the study and thus provided the information that made this report possible.

The NAEP High School Transcript Study project is funded through NCES, in the Institute of Education Sciences of the U.S. Department of Education. The Associate Commissioner for Assessment, Peggy Carr, and the NCES Assessment Division staff—Janis Brown, Steven Gorman, Andrew Kolstad, and Andrew Malizio—worked closely and collegially with the authors to produce this report.

The HSTS project at Westat is directed by Nancy Caldwell, with assistance from Stephen Roey and Carolyn Shettle. In addition, Barbara Brickman, Kay Gallagher, Carol Hannaford, Lloyd Hicks, John Kalter, Yan Yun Liu, Keith Rust, and Neha Singh contributed to the production of this report.

The authors greatly appreciate the comprehensive review of this report by several reviewers at NCES—Janis Brown, Steven Gorman, Andrew Kolstad, Ralph Lee, Andrew Malizio, Jeffrey Owings, Lisa Hudson, and Marilyn Seastrom—and ESSI—John Clement, Linda Hamilton, and Elizabeth Jacinto. The authors also thank Aileen Waters from ESSI for providing administrative assistance.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

List of Appendixes

TABLE OF CONTENTS (continued)

List of Appendixes (continued)

TABLE OF CONTENTS (continued)

List of Tables

TABLE OF CONTENTS (continued)

List of Tables (continued)

TABLE OF CONTENTS (continued)

List of Figures

List of Exhibits

TABLE OF CONTENTS (continued)

List of Exhibits (continued)

FOREWORD

The 2000 High School Transcript Study (HSTS 2000) was conducted by Westat 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' coursetaking patterns in the nation's secondary schools. Since previous transcript studies measured the coursetaking patterns of 1982, 1987, 1990, 1992, 1994, and 1998 graduates, one research objective was to study changes in these patterns. Another research objective was to compare coursetaking patterns to study results on the 2000 National Assessment of Educational Progress (NAEP) mathematics and science assessments. 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 nationwide information about the levels of educational achievement of elementary and secondary school students.

The 2000 High School Transcript Study is documented in three reports:

- *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990–2000*—This summary report highlights major findings from the HSTS 2000 and examines the trends and changes in high school curriculum and student coursetaking patterns for the decade between 1990 and 2000.

- *The 2000 High School Transcript Study User's Guide and Technical Report*—The User's Guide and Technical Report documents the procedures used to collect and summarize the data. It also provides information needed to use all publicly released data files produced by the study.

- *The 2000 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 2000, 1998, 1994, 1990, 1987, and 1982 High School Graduates*—This upcoming report provides extensive tables that summarize the coursetaking patterns of high school students who graduated in 2000 and compare them to those of their counterparts in 1982, 1987, 1990, 1994, and 1998. The report also describes the relationship of the coursetaking patterns of 2000 graduates and the mathematics and science proficiencies as measured by the 2000 National Assessment of Educational Progress.

1. INTRODUCTION

This technical report documents the procedures used to collect and summarize data from the 2000 High School Transcript Study (HSTS 2000). Chapters in the report detail the sampling of schools and students (chapters 2 and 3), data collection procedures (chapter 4), data processing procedures (chapter 5), and weighting procedures (chapter 6). Chapter 7 describes the HSTS 2000 data files and codebooks that are encompassed by this report. Appendix A contains the HSTS 2000 data collection and documentation forms, and appendix B contains the associated NAEP 2000 study questionnaires. Appendix C describes the Classification of Secondary School Courses (CSSC), which was used to code the courses on the HSTS 2000 transcripts, and provides a complete listing of CSSC codes. The codebooks for all of the HSTS 2000 data files may be found in appendixes D through P. Appendix Q is a glossary of terms.

This chapter provides an introduction to the HSTS 2000 through a series of question-and-answer sections, each providing a brief overview of specific aspects of the study. At the end of each section, the reader is directed to a subsequent chapter or chapters in this report, or to the companion report *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990–2000* (Perkins et al. 2004), where selected topics are discussed in greater detail.

■ What is the High School Transcript Study?

Over the years, various reform efforts have sought to improve the quality of education across the United States. In the early 1980s, the focus was on statewide curricula in core courses, a response to the watershed report, *A Nation at Risk* (National Commission on Excellence in Education 1983). Since then, national efforts have addressed several issues concerning quality education, analyzing the content of courses in specific subject areas (mathematics and science, for example), the number of courses completed, and when courses are completed.

The High School Transcript Study (HSTS) is a periodic survey that provides educational professionals, such as administrators, policymakers, and researchers, with information regarding curricula being offered in our nation's high schools and the coursetaking patterns of high school students. It can also be used to provide information on the relationship of student coursetaking patterns to achievement as measured by the National Assessment of Educational Progress (NAEP). NAEP is an ongoing, periodic assessment of educational achievement in U.S. schools.

The transcript studies serve as a barometer for changes in high school student coursetaking patterns. School course offerings and student coursetaking patterns provide valuable information about the rigor of high school curricula across the nation. The first national transcript study was conducted by NCES in 1982 and captured baseline information on high school students' patterns prior to the publication of *A Nation at Risk* and the resulting changes in curricula and educational reform.

For HSTS 2000, about 21,000 transcripts of students who graduated from public and nonpublic high schools were collected from a nationally representative sample of schools from May through October 2000. The survey was conducted in conjunction with the 2000 National Assessment of Educational Progress (NAEP 2000) mathematics and science assessments in the 12th grade. A description of this survey can be found on the NAEP home page at http://nces.ed.gov/nationsreportcard/.

Since similar studies were conducted on the coursetaking patterns of graduates over the years, changes in these patterns can be studied and compared. Table 1 lists the seven studies that have been conducted beginning in 1982 involving the collection of transcripts of high school graduates.

Table 1. High school transcript studies: 1982–2000

Study	Approximate number of transcripts collected
1982 High School and Beyond	12,000
1987 High School Transcript Study	25,000
1990 High School Transcript Study	21,000
National Education Longitudinal Study of 1988 Second Follow-Up (1992)	17,000
1994 High School Transcript Study	25,000
1998 High School Transcript Study	25,000
2000 High School Transcript Study	21,000

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School and Beyond (HS&B), 1982; National Education Longitudinal Study of 1988 (NELS:88) Second Follow-Up, 1992; High School Transcript Study (HSTS), Selected years, 1987-2000.

- **Was participation in HSTS 2000 voluntary? Are the data confidential? Are student names or other identifiers available?**

Students' transcripts were collected by field workers for the sample of students that were selected for the NAEP 2000 assessment. Schools were contacted regarding whether or not to inform parents or obtain parental consent. Generally, schools do not require parental or student notification or consent for the HSTS because there is no burden placed on the student. However, if a school requires that

students and/or parents be notified, or that their consent must be obtained, that request is met. It should be noted that, in the history of these studies, no school has ever requested consent forms for participation.

The data obtained from the transcript study were kept strictly confidential. Student names and any other identifiable information were deleted from the copies of the transcripts before these materials left the schools. Furthermore, in schools that participated in the NAEP assessments, each student received a NAEP ID that was also used in the HSTS. The list that linked the student's name with that NAEP ID remained in the school. High School Transcript Study staff did not have access to that list and could not recreate it if it were lost.

The restricted-use HSTS 2000 data files do not contain the students' names or other variables that directly identify the sampled students. Data files do contain the students' NAEP ID, which enables researchers to link the transcript data to the NAEP data. The HSTS follows NCES' strict procedures regarding the confidentiality of data files.

For more information regarding how the student transcripts were obtained for the study, please refer to chapter 4. For detailed information on how to obtain the restricted-use data files, and a description of the files, please see chapter 7.

- **What contextual background data does HSTS 2000 provide?**

Contextual background data for the HSTS 2000 are obtained from the NAEP 2000 questionnaires, the high school transcripts, and various school-level forms completed by a school coordinator or counselor.

QUESTIONNAIRES

- *School Questionnaire*: The School Questionnaire (see appendix B) was a 54-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 2000 for the NAEP participating schools. Schools that did not participate in NAEP 2000 were also asked to complete the questionnaire.

- *Students with Disabilities/Limited English Proficiency (SD/LEP) Questionnaire*:[1] Prior to 1996, the questionnaire that collected information from school staff about

[1] LEP is used both to identify a specific skill level with regard to English proficiency and, more broadly, to refer to all students for whom English proficiency is an issue.

students with disabilities and students with limited English proficiency was called the Individualized Education Plan/Limited English Proficiency (IEP/LEP) Questionnaire. It was retitled the SD/LEP Questionnaire in 1996. The SD/LEP Questionnaire was completed for students sampled for NAEP and identified by the school as having a disability and/or limited English proficiency. Schools were asked to have the person most knowledgeable about the student complete the 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 information collected in this questionnaire can be found on pages 20–21.

For schools participating in NAEP 2000, the SD/LEP Questionnaires were collected as part of the NAEP procedures. Questions 1 and 2 were used to determine which section(s) of the questionnaire should be completed. Part A (questions 3 through 19) was answered for a student with a disability. Part B of the questionnaire (questions 20 through 41) was completed for an LEP student. If a student was classified as both SD and LEP, the entire questionnaire was completed. A copy of the questionnaire is included in appendix B. SD/LEP Questionnaires were also collected from schools that did not participate in NAEP by field staff involved with the HSTS 2000 data collection.

TRANSCRIPTS

The student transcripts provided data that were coded and entered into the data system by trained coders. These data included the following:

- Date student enrolled in high school;
- Date student graduated;
- Rank in class;
- Size of class;
- Grade Point Average (GPA);
- Days absent each year;
- Standardized test scores and honors (where available);
- List of courses taken in high school, including the grades received and the number of credits earned for each course; and
- Total number of credits received and, in many cases, total number of credits attempted.

SCHOOL FORMS

- *Transcript Request Form (TRF)*: A field worker completed a Transcript Request Form upon returning to a school to obtain requested student transcripts. The form contained student demographic data, including Title 1 and National School Lunch Program participation status, as well as their graduation status.

- *School Information Form (SIF)*: The completed School Information Form contained information about the school in general, such as sources of data collection information within the school, course description materials, graduation requirements, and grading practices.

- *School-level Catalog or Course Lists*: Data entry personnel entered a list of all course titles appearing in the catalogs provided by the schools. A curriculum specialist selected which course titles to enter, and a concerted effort was made to standardize the format of titles.

For more information, please refer to chapters 4 and 5.

- **What were the HSTS 2000 procedures for collecting data?**

The field workers for the HSTS 2000 were drawn from the pool of NAEP field supervisors and were trained in the data collection procedures. Eligible schools participating in NAEP were informed about the HSTS 2000 when they received information about NAEP 2000. This information included procedures that would be used to ensure confidentiality of the data, and the amount and nature of school staff time required for HSTS 2000 participation. Whenever possible, HSTS field staff assisted the school staff with data collection.

For eligible schools that agreed to cooperate, students sampled for NAEP 2000 were included in the HSTS 2000 sample. A brightly-colored Disclosure Notice (see exhibit A-1 in appendix A) was placed in their folder both to alert school personnel that information contained in the student's folder would be used for the HSTS 2000, and to serve as a visible marker for identifying the folders of selected students to facilitate finding their transcripts later.

Initial HSTS 2000 information requested from schools and collected by field workers at the time of the NAEP 2000 assessment included information which they were asked to provide on the School Information Form (SIF) (see exhibit A-2 in appendix A). Other requested information included copies of their school's course catalogs for the four most recent school years, including the current 1999-2000 school year, and three sample transcripts. Information provided on the SIF included the appropriate date

for the field workers to return to the school and obtain the transcripts. When completing the SIF, field workers also gathered general school policy data, including information about class periods, credits, and graduation requirements. This school policy data may appear in either the course catalog or a separate document.

Field workers completed checklists for the materials they obtained. These checklists served two purposes:

1. They guided field workers in obtaining materials with the maximum amount of information possible that would be useful in the HSTS 2000.

2. They provided HSTS 2000 staff with a quick way to review the materials, so that they could request additional information if needed.

This information was collected during visits to the schools prior to and at the time of the assessment.

When graduation information was posted on transcripts, a field worker returned to the school to obtain the requested transcripts. Schools that stored their transcripts electronically could provide an electronic copy of their transcripts. For schools that kept paper copies of their transcripts, the transcripts were manually pulled from their folders and photocopied at the school. The Disclosure Notice placed in students' folders at the time of the NAEP 2000 assessment helped to facilitate transcript collection in participating NAEP schools.

Once the transcripts were provided, the field worker completed the Transcript Request Form (TRF) (see exhibits A-3 and A-4 in appendix A). The worker first reviewed the transcripts to ensure that a transcript was received for each 12^{th}-grade student who was selected for the NAEP 2000 assessment, whether or not that student had graduated. (Nongraduates were removed from the files at a later stage.) Which transcripts were received and not received were recorded on the TRF. For received transcripts, the field worker also recorded the sampled students' names and school exit status on the TRF, along with any missing student demographic information.

The field worker then checked each transcript for eligibility, understandability (e.g., whether all the codes on it were defined on the transcript or explained in the SIF), and completeness. He or she labeled each transcript with preprinted labels containing the School ID and the NAEP ID for the student. For students with missing transcript information, the field worker completed a Documentation of Missing Transcripts form (exhibit A-6) to explain any omissions.

After the field worker reviewed the transcripts for completeness and accuracy, he or she prepared the transcripts for removal from the school. 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.

For schools that did not participate in NAEP 2000 but agreed to take part in the HSTS 2000, contact was made near the end of the 1999-2000 school year, once the students' final data were posted on their transcripts. The same information obtained for schools participating in NAEP was collected for schools not participating in NAEP. Depending upon the number of students in the graduating class, up to 50 students were randomly selected from the class list to participate in the transcript study.

For more information, and a detailed description of the process used in obtaining materials for the HSTS 2000, please refer to chapter 4.

- **What is the Transcript Request Form (TRF)? How is the TRF obtained and what information does it contain?**

When a field worker returned to the school to obtain the requested transcripts, the worker brought a Transcript Request Form (TRF) on which to record information about the HSTS sampled students. There were two versions of the TRF, Version 1 (exhibit A-3) and Version 2 (exhibit A-4). For each NAEP school, the field worker was given a TRF Version 1. Data available from NAEP 2000 files (NAEP ID and demographic variables) were already preprinted on the form. This information included the student's NAEP ID, gender, birth month and year, race/ethnicity, SD status, LEP status, receipt of Title 1 services, and National School Lunch Program participation. The field worker recorded the student's name, school exit status, and whether or not a transcript was received for the student.

The completed TRFs contained the following information:

- *Student Name* – Since names were never removed from the school, this column was blank when the TRFs were printed. The field worker first recorded the first name, middle initial, and last name of each assessed, absent, or excluded student listed on the NAEP 2000 Administration Schedule (exhibit A-5). The names were recorded only to ensure that the correct student folders were used.

- *NAEP ID* – The 10-digit NAEP assessment booklet numbers, or the SD/LEP questionnaire numbers for students excluded from the 2000 assessment, were preprinted on the TRF in ID order. This column on the TRF identified all students for whom transcripts were needed.

- *Exit Status* – Using information provided by the school, field workers assigned a code to describe each student's outcome at the school. The Exit Status codes are listed and defined on page 18.

- *Birth Date, Gender, and Race/Ethnicity* – Demographic information was generally preprinted for each sampled student. If not preprinted, it was recorded from the NAEP 2000 Administration Schedule. If the school informed a field worker that some of this information was incorrect, the field worker entered the correct information on the TRF.

- *SD and LEP Status* – For each student, it was recorded whether or not the student was classified by the school as SD and/or LEP.

- *National School Lunch Program (NSLP) and Title 1* – Field workers recorded either "Yes" or "No" for student participation in each of these programs.

- *Transcript Received* – Field workers checked this column to document that the transcript for a given student had been received.

For each non-NAEP participating school, the field worker was given a TRF Version 2. This form captured the same data as Version 1 with the exception of a NAEP ID. Students from non-NAEP schools were given unique 10-digit IDs with 990 prefixes.

Personal identifiers were also removed from the Transcript Request Forms. Before sending the TRFs from the school, the field worker cut off the portion that contained the students' names to comply with confidentiality provisions. The portion with the names was left in the school's NAEP folder.

For further information, please refer to chapter 4.

- **What is a course catalog? What are the various types of course catalogs? How are the course catalogs obtained?**

A course catalog is a listing and description of courses a high school offers. High schools generally publish a course catalog each year. A Classification of Secondary School Courses (CSSC) code (see the following section or chapter 5) was assigned to individual courses listed in a school's course catalog, based on the descriptions the catalog provided. The coded course catalogs were then used to assign CSSC codes to individual course titles listed on the student transcripts. The course catalogs also formed the basis for the HSTS 2000 course offerings data file.

The HSTS has identified five types of course catalogs. Ranked from highest to lowest in terms of usefulness for catalog coding, the five catalog types are as follows:

1. A school-level catalog providing course titles and descriptions;

2. A district-level catalog which indicates which courses were offered at the HSTS participating school;

3. A course list by department that includes general descriptions of course offerings by department;

4. A school-level course list without descriptions; or

5. A district-level catalog that does not indicate which courses were offered at the HSTS participating school.

The highest-level catalog available is used for catalog coding.

Field workers requested course catalogs when they first contacted a school, then collected and carefully reviewed them when they visited the school for sampling. Field workers verified that the catalog contained all the courses that the 12th-graders of that year took in that school, including vocational, remedial, honors, special education, off-campus courses, or courses taught in a language other than English. If any course listings were not in the catalog, every effort was made to obtain additional information from school personnel to document the existence of such courses and to describe them. After that review, the course catalogs were taken from the school.

In most cases, the current course catalog and the ones from the three preceding years were collected. This collection allowed tracking of any changes in course offerings or the curriculum in the four years the sampled students attended high school. It also allowed the catalog coders to review any course title on the transcript and accurately match it to a description in the catalog, even if the curriculum or the course titles had changed during those four years.

For further details, please refer to chapter 4.

- **What is a Classification of Secondary School Courses (CSSC) code and how is it used? Are there any other coding systems that are being used in similar studies?**

To compare transcripts from different schools, it was necessary to code each of the courses entered from the transcripts using a common course coding system. The coding system employed for this purpose was a modification of the system presented in *A Classification of Secondary School Courses* (Ludwig et al. 1982). The CSSC, which contains 2,268 course codes, is a modification of the college course classification system presented in *Classification of Instructional Programs* (Morgan, Hunt, and

Carpenter 1991). Both course coding systems use a three-level, six-digit system for classifying courses. The CSSC uses the same first two levels as the Classification of Instructional Programs (CIP), which is represented by the first four digits of each code.[2] The third level of the CSSC (the fifth and sixth digits of the course code) is unique to the CSSC and represents specific high school courses.

A taxonomy of course subject areas was developed for the 1987 High School Transcript Study. This taxonomy, documented in the 1987 HSTS tabulations (Thorne 1988), was developed with an emphasis towards academic courses. Computer-related courses were considered as a separate non-vocational subject, and there were fewer subgroups defined for vocational and personal courses. This taxonomy was applied to data from the 1982 High School and Beyond (HS&B) First Follow-Up Study and the HSTS 1987 data. The 1990 High School Transcript Study used a slightly expanded version of the same taxonomy in its reports.[3]

Starting with the 1994 study, the HSTS switched over to the Secondary School Taxonomy (SST). Originally developed in the late 1980s by the National Assessment of Vocational Education,[4] the SST has a less purely academic emphasis and a more richly defined group of vocational education categories than the taxonomy developed for the earlier HS&B and HSTS studies. Computer-related courses became vocational courses, and general skills and military science courses became new subject areas. So to maintain comparability with the earlier transcript studies, the 1987 and 1990 HSTS studies, along with the 1982 HS&B study, were recoded using the SST.

With 2,268 codes in the CSSC, it is often neither practical nor desirable to tabulate estimates of each possible CSSC code. It is typically more useful, however, to analyze the courses in larger subject areas such as English, social studies, mathematics, or science. There is also interest in subgroups of these subject areas, such as biology, chemistry, and physics. The taxonomy presented in appendix C provides the structure for aggregating the courses to subject areas. For those researchers interested in the occurrence of each CSSC code among the NAEP-related high school transcript studies, the forthcoming online publication *The 2000 High School Transcript Study Tabulations: Comparative Data on Credits*

[2] Specifically, 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 students with disabilities.

[3] The 1990 study added 18 new codes to the CSSC and to the taxonomy. The full taxonomy is documented in both *The 1990 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1990, 1987, and 1982 High School Graduates* (Legum et al. 1993a) and *USER'S MANUAL: 1990 High School Transcript Study* (Legum et al. 1993c).

[4] A description of the development of the SST is provided in *The Secondary School Taxonomy Final Report* (Gifford, Hoachlander, and Tuma 1994).

Earned and Demographics for 2000, 1998, 1994, 1990, 1987, and 1982 High School Graduates (Perkins, Roey, and Brown forthcoming) will contain a table that provides estimates for each CSSC code that appears in the HSTS student transcripts.

For further details about the CSSC, please refer to chapter 5. For a list of CSSC codes used in HSTS 2000 catalog and transcript coding, please refer to appendix C.

- **How are codes added to the CSSC? Are they ever deleted?**

Codes are added to the CSSC whenever courses are found in the catalogs that have no match in the CSSC. Highly trained coders were used to code the school catalogs received from the field workers. These coders browsed through the catalogs and matched the appropriate CSSC codes to the courses offered, according to the content and description of the course. If a course that was offered did not have a matching CSSC code in the existing list, the coders wrote that course description in a special suggestion list. After the catalogs were reviewed, and all but those courses on the suggestion list were coded, a Coding Specialist reviewed the suggestion list and tried to match these courses to existing CSSC codes. If a course did not have a matching CSSC code, a new CSSC code was generated.

The high school curriculum may change each year or every few years. New courses are added, old courses are taken out of the curriculum, and some courses are combined with others to produce new courses. For every High School Transcript Study, the need arises to examine the list of CSSC codes and decide whether each of the courses that were offered in that particular year has a matching CSSC code that can adequately describe it. The CSSC code list contains 2,268 codes and descriptions of courses offered by high schools nationwide. In 1994, 12 new CSSC codes were added to the list. In 1998, the CSSC's computer science curriculum changed dramatically. New courses such as Web Design, Java Programming, and C++ Programming were added. Many courses that were labeled as honor courses in the past were reclassified as Advanced Placement (AP) courses. Many International Baccalaureate (IB) courses were added as well. In all, a total of 83 new or revised codes were added to the CSSC in 1998. In 2000, two CSSC codes were added, one in science and one in computer-related studies.

The examination of CSSC codes in HSTS 2000 also revealed five CSSC codes that were either duplicate codes or previously added codes that have never been used. These five CSSC codes were eliminated from the HSTS 2000 master CSSC list.

For further information about the CSSC codes, please refer to chapter 5. For a list of CSSC codes used in HSTS 2000 catalog and transcript coding, please refer to appendix C.

- **How are the catalogs coded? What special requirements are needed from the coders? How are they trained?**

The staff hired to code the school catalogs consisted of individuals who had an extensive background in education, mostly teachers and counselors familiar with school curricula and the education system. These staff members underwent training to familiarize themselves with the CSSC coding scheme and how to code a course based on available catalog information. For several days, they were given exercises and tasks to ensure that they could code a course title with the appropriate CSSC code.

To ensure consistency and quality, catalog coding decisions were based 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. The coder 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 the Computer Assisted Coding and Editing (CACE) system, the coder looked at each course catalog title on the screen, located it in the hardcopy catalog, and reviewed whatever description was available. The coder then selected the most appropriate CSSC code for the course. Wherever possible, the catalog coder selected codes based on a course description rather than on the title. All of the courses found in the catalogs were coded months prior to the receipt of the student transcripts.

After selecting the CSSC code, the coder reviewed the course sequence, off-campus, language, remedial, honors, and special education status flags for that course and edited them as needed. If the coder found courses in the CACE catalog listing that should not be there, the courses were 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 the special education courses.

For the specific steps of the coding procedure please refer to chapter 5.

- **How are the HSTS 2000 data entered?**

The data from the HSTS 2000 were processed along three simultaneous paths as follows:

1. The process of sampling student information;
2. The Computer Assisted Data Entry (CADE) system; and
3. The Computer Assisted Coding and Editing (CACE) system.

With the exception of the transcripts and the course catalogs, some data entered for each process were collected by field personnel and some data had already been assembled for NAEP 2000 into data files by the Educational Testing Service (ETS). The relevant NAEP 2000 data files were obtained from ETS and merged with the HSTS 2000 data collected from non-NAEP 2000 participating schools. Appropriate checks were made to ensure that only one set of data was entered for a school or a student, and procedures were developed to resolve inconsistencies among the data sources.

When entering and cleaning the data for the study, the following tasks were performed:

- Establishing student ID control lists;
- Entering transcript data;
- Coding course catalogs;
- Matching transcript course titles to catalog titles;
- Standardizing credits and grades; and
- Performing quality control checks.

These steps involved the entry and coding of the students' transcripts and the schools' course catalogs, as well as matching the courses on the coded catalogs to the courses on the transcripts.

Each of these steps is described in detail in sections 5.1 through 5.6 of chapter 5.

- **How is the HSTS 2000 related to the 2000 National Assessment of Educational Progress?**

The HSTS is conducted in conjunction with the National Assessment of Educational Progress (NAEP). The HSTS 2000 was designed to allow an analysis of the coursetaking patterns of

students who graduated from American public and nonpublic high schools in 2000. It was further designed so that data on students' coursetaking patterns can be linked to the NAEP 2000 assessment results. NAEP provides results about subject matter achievement, instructional experiences, and school environment, and reports these results for populations of students (e.g., 12th-graders) and selected subgroups of those populations (e.g., male students). Changes in the relationship of HSTS coursetaking to NAEP performance can also be examined for similar studies in 1994 and 1998.[5]

NAEP provides the HSTS with data on assessments in different subjects. For HSTS 2000, the proficiency estimates of mathematics and science were provided.

For a comprehensive description of the HSTS and NAEP, please refer to chapter 2.

- **How are the samples of schools and students in NAEP 2000 related to the HSTS 2000 samples?**

To maintain as many links as possible with NAEP 2000 scores, schools refusing to participate in NAEP 2000 were replaced by substitute schools, and the substitute schools, not the refusals, were asked to participate in the HSTS 2000. Of the 359 eligible schools in the original NAEP sample, 277 original/substitute schools participated in the HSTS 2000 survey, of which 265 were originally sampled and 12 were substitute schools. Of the 277 participating schools, 248 schools cooperated with both HSTS 2000 and NAEP 2000 and the links for the students were maintained, 13 schools cooperated with HSTS 2000 and NAEP 2000 but the links for the students were not maintained, and 16 schools cooperated with HSTS 2000 but not with NAEP 2000. The links between the students and their IDs are maintained at the schools in order to preserve the confidentiality of the students. As there is an interval of around six months from the time the student is assessed and the time the transcripts are collected, some schools inadvertently destroyed these lists. It was not possible to reconstruct these lists.

A total of 23,440 students were selected for the HSTS 2000. Of these, 22,010 students were from schools that maintained their NAEP 2000 administration schedules and were identified by their NAEP booklet numbers. Another 630 students were from schools that participated in NAEP 2000 but had lost the link between student names and NAEP booklet numbers, and 800 were from schools that did not participate in NAEP 2000.

[5] The 1994 and 1998 transcript data were collected by Westat in coordination with the 1994 and 1998 NAEP (Legum et al.. 1997; Roey et al.. 2001b).

Because sampling was performed in most schools prior to graduation, not all sampled students were, in fact, graduates. Only graduates, however, were eligible for inclusion in the transcript study. It was determined that, of the 23,440 students in the sample, 21,085 actually graduated by October 2000 and 2,355 did not. From the 21,085 graduates, 20,931 transcripts were collected and processed, while no transcripts were received from 154 graduates.

For further information regarding this topic, please refer to chapter 3.

Can the HSTS 2000 results be compared to other transcript studies?

Between 1982 and 2000, the National Center for Education Statistics (NCES) conducted seven high school transcript studies associated with the High School and Beyond (HS&B) survey in 1982, the Second Follow-Up to the National Educational Longitudinal Study in 1988 (NELS:88), and the National Assessment of Educational Progress (NAEP) in 1987, 1990, 1994, 1998, and 2000. One research objective of NAEP HSTS 2000 was to study changes in the coursetaking patterns among high school students over time, comparing its results with the other NCES-conducted high school transcript studies. While results are reported for trends over time, it should be noted that some differences exist between the high school transcript studies and some direct comparisons are cautioned.

The first high school transcript study was conducted in 1982. The 1982 study was part of the first follow-up to the longitudinal HS&B study. Transcripts were collected from seniors who were members of the 1980 HS&B sophomore cohort. In 1987, the first transcript study that was associated with the NAEP was conducted. The results from the NAEP HSTS 1987 were used to compare coursetaking patterns of high school graduates in 1982 and 1987. The four subsequent NAEP HSTS studies in 1990, 1994, 1998, and 2000 have been used by NAEP to track changes in the coursetaking patterns of high school graduates. For researchers interested in a data point between the NAEP HSTS studies in 1990 and 1994, the transcript component of the second follow-up to NELS:88 may be used. Numerous NCES studies and reports have included transcript data from the NELS:88 second follow-up study for comparisons with the results from the other transcript studies.

For more information about comparisons among the different HS&B and HSTS studies, please refer to chapter 1 of *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990-2000* (Perkins et al. 2004) (also referred to as the Summary Report).[6] For discussion

[6] The Summary Report can also be found at http://nces.ed.gov/nationsreportcard/.

about comparisons with the transcript component of the Second Follow-Up to NELS:88, please refer to Appendix A of *National Education Longitudinal Study of 1988, Second Follow-Up: Transcript Component Data File User's Manual* (Ingels et al. 1995). The similarities and differences between the high school transcript studies' data (NAEP, NELS, HS&B) are also described extensively in the *NCES Handbook of Survey Methods* (Thurgood et al. 2003). The handbook looks at the comparability of the high school transcript studies' data based upon five criteria: (1) sample sizes; (2) oversampling of subgroups; (3) eligibility criteria for inclusion into the studies; (4) representativeness of cross-sectional and longitudinal populations; and (5) coding differences.

- **What is a weight and how is it determined?**

A weight is a numeric value assigned to a sampled item (e.g., school or student) so that the sample can reflect the entire population that it measures. The HSTS 2000 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 design requires that the user of the HSTS 2000 data use sampling weights to ensure valid analysis of the transcript data.

Several sets of weights were created for HSTS 2000. The nonlinked weights, also called the "student weights," were assigned for all eligible sampled students with completed, missing, or unusable transcripts in the transcript study. "Eligible" students are students who graduated in 2000, and "unusable" transcripts were those transcripts with less than 75 percent of the credits required by the school to graduate. Weights were set to zero for missing and unusable transcripts.

Weights were also created for students that were sampled for NAEP, whether or not these students had participated in NAEP. These weights are referred to as "linked" weights since the students were part of the NAEP study. Weights were assigned for both assessed and excluded students who graduated and for which usable transcripts were obtained. For the HSTS 2000, two sets of "linked" weights were created. In one set of weights, students with a disability or limited English proficiency students without accommodations were excluded; in the other set of weights, they were included. Since students in NAEP were assigned an assessment of a particular subject (mathematics or science), separate weights were developed for the students in each subject-specific assessment.

Student transcript data were weighted for the purpose of making estimates of coursetaking by high school graduates nationwide. The weights reflected the probability sampling scheme used to arrive at the sample of students for whom transcripts were requested. The final weight attached to an individual student record reflected two major aspects of the sample design and the population being surveyed. The first component, the base weight, was used to expand sample results to represent the total population and reflected the probability of selection in the sample. The second component, the adjustment of the base weight to account for nonresponse within the sample, was implemented to ensure that the resulting survey estimates of certain characteristics (race/ethnicity, size of community, and region) conformed to those estimates known reliably from external sources.

The HSTS 2000 weights were constructed without regard to the NAEP 2000 participation/nonparticipation status of schools and students. The HSTS 2000 weights reflected the impact of sample nonresponse at the school and the student level, and made weight adjustments to decrease the potential bias that might arise through differential nonresponse across population subgroups. Improvements to the precision of weighted estimates also resulted from the application of poststratification factors to the HSTS 2000 weights.

For further information, please refer to chapter 6.

- **Why are there two general sets of weights (linked and nonlinked weights) for HSTS 2000?**

Because the sample of students that participated in both HSTS 2000 and NAEP 2000 assessments was a subset of the larger HSTS 2000 student sample, the students represented in the linked weights databases required a different set of sampling weights. In particular, the school and student nonresponse adjustments will be larger for the linked weights than for the nonlinked weights. These larger adjustments are because a student or school had to participate in both the NAEP 2000 and the HSTS 2000 surveys to qualify as a "respondent" for the linked database. This criterion reduced the number of school and student responses, thereby increasing the nonresponse adjustment factors.

The nonlinked weights, found in the HSTS 2000 student file, allow making generalizations about the graduating 12th-grade population in the year 2000 based on the full sample of HSTS 2000 students for whom transcripts were collected. The linked weights, found in the four HSTS 2000 linked weight data files, allow making generalizations about the graduating 12th-grade population in the year 2000 based on the sample of students for whom both transcripts and NAEP assessment scores were

collected. All HSTS 2000 analyses that involve NAEP 2000 assessment scores should use the appropriate linked weights, while all other HSTS 2000 analyses should use the nonlinked weights.

For more information about the linked weights, please refer to chapter 3.

- **What is the Primary Sampling Unit (PSU)?**

For the HSTS 2000, the primary sampling unit, or PSU, is either a county or group of counties that formed the first-stage sampling units in the HSTS multistage sample. One purpose of the HSTS 2000 was to gather data that could be linked to NAEP results on a nationally representative sample of students who graduated from public and nonpublic high schools in the United States in 2000. For the HSTS 2000 sample of students to be as representative as possible, it included the sampled schools with 12^{th} grades that were selected for NAEP 2000, regardless of whether they participated in the NAEP 2000 assessment.

For further information, please refer to chapters 2 and 3.

- **What is an Exit Status and how it is used?**

The Exit Status is a code that describes the type of diploma the student received. Using information provided by the school, field workers assigned one of the following codes to describe each student's outcome at the school.

1. Graduated with a standard diploma;
2. Graduated with an honors diploma;
3. Received a diploma with special education adjustments;
4. Received a certificate of attendance;
5. Still enrolled in this school;
6. Dropped out;
7. Other, such as transferred, General Equivalency Diploma (GED), or unknown;
8. Out of Scope (i.e., did not meet the eligibility requirements for participation in this study); or
9. Completed course requirements but did not pass required tests for graduation.

In some cases, the Exit Status was determined directly from the transcripts, and sometimes it was provided by other sources at the school. The Exit Status was recorded on the Transcript Request Form and later used to verify that the student indeed graduated and that his/her transcript was eligible for the study. It also provided information about whether or not to include the transcript in the tabulation process. In a few cases, it was determined that a student had not actually graduated and the Exit Status was revised accordingly.

For more information about the Exit Status, please refer to chapters 4 and 5.

- **How are the high school transcripts coded?**

Transcript coding starts with the schools' course catalogs. Course titles appearing in each school's course catalog were keyed into the Computer Assisted Coding and Editing (CACE) system. The resulting list was then checked, verified, and revised as necessary by a catalog coder and supervisor. Then, using CACE, the catalog coder assigned a Classification of Secondary School Courses (CSSC) code to each course listed in the catalog, referring to the catalog itself for a course description. CACE also prompted the catalog coder to set all flags that may pertain to a course, such as those for honors, remedial, or off-campus courses.

Next, using another portion of the CACE system, the catalog coder matched each unique course title appearing on a transcript from a school to a title included in the course catalog from that school. CACE then assigned the linking catalog identification to the transcript course title from that school. For schools that did not provide catalogs or course lists, the transcript courses were title-matched with a "generic" course catalog. The generic catalog included all of the current courses found in the CSSC. Grades and credits were entered for each course in the transcripts and standardized into a consistent system.

In the 1987, 1990, 1994, 1998, and 2000 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 who were taught in a language other than English. In 1998 and 2000, courses offered as Advanced Placement or International Baccalaureate courses were coded separately from other honors-level courses, using both new CSSC codes and new flag values. In addition to codes for Advanced Placement and International Baccalaureate courses, most new codes reflect changes in course offerings in the technology area.

More detailed information about transcript coding, including coder training, is included in chapter 5.

- **What student information is obtained?**

Information gathered for all students included the following:

- gender;
- race/ethnicity;
- birth year;
- birth month;
- student exit status;
- graduation date;
- type of diploma;
- disability status;
- limited English proficiency status;
- whether or not received Title 1 services;
- whether or not participated in the National School Lunch Program;
- date of entry to the school;
- number of days absent in each of four years (9^{th} grade, 10^{th} grade, 11^{th} grade, and 12^{th} grade);
- grade point average; and
- class rank.

In addition, all awards and scores on certain standardized tests (e.g., PSAT, SAT, ACT) taken by each student as reflected on the transcript were listed.

School personnel provided additional information for disabled and limited English proficient students though the NAEP 2000 SD/LEP questionnaire. Additional information collected for disabled students included the following:

- grade-level equivalent performance in English and mathematics;
- proportion of time the student was placed in mainstream and special education classes;
- type and severity of disability; and
- type of accommodation(s) provided for the student.

Additional information collected for students with limited English proficiency included the following:

- English and mathematics grade levels;
- percentage of the day spent in special language programs;
- native language;
- type of specialized instruction;
- the type of accommodation(s) provided for the student in testing; and
- the student's ability to speak, understand, read, and write English.

Chapter 4 discusses the collection of student data.

- **What data files are available for HSTS 2000?**

Table 2 lists the 13 data files that are available on the HSTS 2000 restricted-use data sets.

Table 2. High school transcript study files: 2000

Data File	Description of data file	Number of records on data file
Master CSSC File	Lists the Classification of Secondary School Courses (CSSC), including all modifications made to the original (1982) CSSC during the 1987, 1990, 1994, 1998, and 2000 transcript studies	2,268
Course Offerings File	Provides a listing of the courses offered in the schools included in the study, along with associated CSSC codes	68,238
School File	Provides detailed information on the schools from which the students were sampled	277
Student File	Provides demographic information on all students in the study, as well as sampling weights and summaries of their coursetaking histories	23,522
Mathematics R2 Linked Weights File	Provide weights for use when performing analyses relating transcript data to NAEP 2000 mathematics assessment results (nonaccommodations).	8,941
Mathematics R3 Linked Weights File	Provide weights for use when performing analyses relating transcript data to NAEP 2000 mathematics assessment results (accommodations).	8,998
Science R2 Linked Weights File	Provide weights for use when performing analyses relating transcript data to NAEP 2000 science assessment results (non-accommodations).	11,120
Science R3 Linked Weights File	Provide weights for use when performing analyses relating transcript data to NAEP 2000 science assessment results (accommodations).	11,136
NAEP 2000 Mathematics Assessment Data File	Contains proficiency estimates for each HSTS sampled student who completed the NAEP 2000 mathematics assessment	6,542
NAEP 2000 Science Assessment Data File	Contains proficiency estimates for each HSTS sampled student who completed the NAEP 2000 science assessment	7,982
Tests and Honors File	Provides a list of honors and standardized test results that were included on the transcripts	19,381
Transcript File	Provides a complete list of all courses appearing on the transcripts of students in the study	995,035
SD/LEP File	Provides detailed information on students with disabilities and/or limited English proficiency	2,561

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

The NAEP 2000 assessment data files contain NAEP 2000 scores for the total number of graduates who participated in both the specific NAEP assessment and the transcript study. However, students who did not meet the graduation requirements were later excluded from the transcript study. Their data are present only in the NAEP 2000 assessment files and not in the transcript data files.

2. BACKGROUND: SAMPLE DESIGN

This chapter describes aspects of the NAEP 2000 sample design that affect the HSTS 2000. The HSTS 2000 used all public schools and about a 10 percent subsample of nonpublic schools from the 12th-grade NAEP 2000 assessment. The HSTS 2000 student sample consisted of the NAEP 2000 student sample in these subsampled schools. The focus of chapter 3 is on aspects of the selection of primary sampling units, schools, and students that are specific to the HSTS 2000.

2.1 NAEP 2000 12th-Grade Sample Design

The 12th-grade sample for the 2000 National Assessment of Educational Progress was a multistage probability-based sample of students. This was a national sample in which counties or groups of counties were the first-stage sampling units, and elementary and secondary schools were the second-stage units. The third stage of sampling consisted of the assignment of session type and sample type to sampled schools. The session type refers to the subject(s) being assessed, while the sample type refers to the specific criteria for inclusion that were applied to the session (see section 2.4 for a discussion of the inclusion criteria). The fourth stage involved selection of students within schools and their assignment to session types.

A total of 94 primary sampling units (PSUs) were included in the sample, and a sample of 642 schools actually participated in the assessment for the 12th grade. Various blocks or packages of exercises were administered to students in these schools.

2.2 Selection of NAEP Primary Sampling Units

In the first stage of sampling, the United States—the 50 states and the District of Columbia—was divided into geographic primary sampling units (PSUs). 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). A PSU consists of a Consolidated Metropolitan Statistical Area (CMSA), a metropolitan statistical area (MSA), a New England County Metropolitan Area (NECMA), a county, or a group of contiguous counties in the U.S. (including Alaska, Hawaii, and the District of Columbia). Each

PSU was contained entirely within one of the four geographic regions defined in table 3. 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.

Table 3. NAEP geographic regions used for stratification: 2000

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[1]	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia[1]			Utah
			Washington
			Wyoming

[1]That part of Virginia which 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.
SOURCE: U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, 2000.

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.

The 22 largest PSUs in the United States were included in the PSU sample with certainty. The remaining smaller PSUs were not guaranteed to be selected for the sample. These were grouped into a number of noncertainty strata and one PSU was selected from each stratum. In each region, noncertainty PSUs were classified as MSA (metropolitan) or non-MSA (nonmetropolitan), forming eight major strata. Within each major stratum, further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics, yielding 72 strata. The number of such strata formed within each major stratum is shown in table 4.

Table 4. The number of noncertainty strata in each major stratum for the NAEP national main assessment: 2000

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

SOURCE: U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, 2000.

The strata were defined so that the sum 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, percent of persons age 25 or over with college degrees, percent of persons age 25 or over who completed high school, and the civilian unemployment rate. Up to four of these characteristics were used to define a major stratum. For each major stratum, the characteristics used were chosen by modeling PSU-level State NAEP mean reading proficiency scores for 1988, 1990, and 1992. The same PSU geographic definition is used for the three NAEP studies, as well as for NAEP 2000. A linear regression model was run using the average reading scores for the three years against the various socioeconomic characteristics at the PSU-level. The characteristics that were most correlated with the average reading scores were selected as the stratum variables.

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 PSUs were selected with probability proportional to size (PPS) with the twin aims of obtaining approximately self-weighting samples of students, and having approximately equal workloads in each PSU.

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, PSUs varied considerably as to their probability of selection, since they varied greatly in size. In each region, noncertainty PSUs were classified as either metropolitan (MSA) or nonmetropolitan (non-MSA). The 36 selected noncertainty MSA PSUs had probabilities ranging from 0.03 to 0.58, while the 36 non-MSA PSUs had probabilities ranging from 0.03 to 0.11.

Since one PSU was selected from each noncertainty stratum, the distribution of the noncertainty PSUs in the sample is the same as the noncertainty strata, as shown in table 4.

2.3 Selection of NAEP 2000 Schools

For NAEP 2000, the second-stage of selection was the sampling of schools. A frame of 12th-grade schools was created by combining the NCES 1997–1998 Common Core of Data (CCD) frame of public schools and the NCES 1997–1998 Private School Universe Survey (PSS) file of nonpublic schools. The sampling frame of eligible 12th-grade schools was restricted to the selected 94 PSUs. There were 6,831 public and 4,272 nonpublic schools on the final school sampling frame.

Public schools from CCD included regular and state-run public schools, Bureau of Indian Affairs (BIA) schools, and Department of Defense Education Activity (DoDEA) schools. Regular and state-run public schools were schools with students who were classified as being in a specific grade, as opposed to schools having only "ungraded" classrooms. These schools included statewide magnet schools and charter schools. Both graded and ungraded schools are included on the CCD, though only graded schools were included on the NAEP school sampling frame.

Nonpublic school information was collected from the PSS conducted by the National Center for Education Statistics. The PSS list of schools is an ongoing registry of nonpublic schools that is updated prior to the survey through two sources. The first source, called the list frame, is a conglomeration of a number of lists from several associations, states, and so on. The second source uses an area frame to identify and represent schools not on the list frame.

For each school in the 12th-grade frame, estimates were made of the number of eligible students in the 12th grade. This estimate was used to determine a school's measure of size for sampling purposes. For the estimated 12th-grade student enrollment, public schools used the average student enrollment per grade (calculated as the total school enrollment from CCD divided by the school's grade range), and nonpublic schools used the reported 12th-grade enrollment from the PSS file.

High-minority public schools on the frame were also identified for oversampling. A school was classified as high-minority if the percentage of Hispanic and Black students was reported to be

greater than 15 percent and the number of Hispanics and Black students was reported to be at least 15. Otherwise the school was classified as low minority.

Schools were selected (without replacement) across all PSUs, systematically from a sorted list with probabilities proportional to assigned measure of size, which was a function of the estimated number of 12th-grade students. The sorting variables included certainty/noncertainty PSU classification, NAEP region, public/nonpublic classification, type of location, high/low minority classification, PSU stratum, school type, and estimated grade enrollment. The order of the sort differed depending on public and nonpublic school classification and certainty/noncertainty PSU classification.

High-minority public schools were given double the probability of selection of a public school not designated 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 sample size, this procedure reduces somewhat the reliability of estimates for all students as a whole and for those students not Black or Hispanic.

In NAEP 2000, nonpublic schools were heavily oversampled to meet explicit target sample sizes for reporting group (Catholic, Lutheran, Conservative Christian, Other Religious, Nonsectarian, and Independent) in order to provide reliable NAEP estimates for such students. The target student sample size was 6,000 for Catholic students and 1,500 each for the other reporting groups. In HSTS 2000, however, the oversampling of nonpublic schools was reversed so that the nonpublic school students in the HSTS were represented in proportion to their prevalence in the general 12th-grade student population (see chapter 3).

The 1997–1998 CCD files do not contain schools that opened between 1998 and the assessment dates. Therefore, special procedures were implemented to be sure that the NAEP 2000 assessment represented students in new public schools. Small school districts—those that contained only one eligible school—were handled differently from large school districts, which contained more than one eligible school. In small school districts, the schools selected 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 12th grade 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 CCD file. 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 the list of eligible schools derived from information on the CCD files. Frames of eligible new schools were then constructed for 12th grade, and samples of new schools were selected systematically with probability proportional to eligible enrollment using the same sampling rates as for the CCD schools. As a result of this process, one new public school was selected.

Potential substitute schools were selected for all sampled schools in the NAEP 2000 where a close match could be identified. In the NAEP 2000, a new procedure was introduced to identify substitutes. No sampled school was assigned more than one substitute, and no school was assigned to be a substitute for more than one school. The criteria for assigning substitutes were quite strict; many sampled schools were not assigned substitutes at all as there were no schools that met the necessary criteria to be a substitute.

Substitutes were assigned by matching on minority composition and estimated number of eligible students. Sampled schools could only have substitute schools in the same school type group, with school type group defined as regular public, Bureau of Indian Affairs, Department of Defense, other public, Catholic, non-Catholic religious, and other nonpublic. Public schools could only have schools in the same PSU and with the same locality type as substitutes. Catholic schools could only have schools in the same district (usually diocese) as substitutes.

A nonparticipating school was replaced by a substitute when the nonparticipating school was considered a final refusal. Of the 642 participating 12th-grade sampled schools, 45 were substitutes.

2.4 Assignment of Sessions and Sample Type to Schools for NAEP

Twelfth-grade schools were assigned two types of sessions, mathematics and science. Schools were assigned either one or two sessions based on the estimated number of grade-eligible students from the frame. It was assigned one session if its estimated grade enrollment was less than 25 students, and two sessions if it was 25 or more students. Schools with two sessions were assigned one of each session type. Schools allocated a single session were systematically assigned a session type of either mathematics or science at rates varying by public/nonpublic schools. For public schools, 7 out of 16

single session schools were assigned mathematics, while 9 out of 16 were assigned science. For nonpublic schools, half of the single session schools were assigned mathematics and the other half science.

To determine the effect of using different criteria for excluding students from the assessment, two different sample types were assigned to schools. In sample type 3 (S3) schools, accommodations were offered to students with disabilities (SD) and students with limited English proficiency (LEP). In sample type 2 (S2) schools, no assessment accommodations were offered to SD/LEP students. Sample type was assigned to schools so that 50 percent of the schools were assigned S2 and 50 percent were assigned S3. Schools that were sampled for more than one grade were assigned only one sample type, which was used for all sampled grades.

2.5 NAEP Student Sampling

The fourth stage of sampling for NAEP 2000 involved the selection of students within the sampled schools. The student samples included oversampling of Black and Hispanic students in low minority public schools and disabled and/or limited English proficiency (SD/LEP) students in all schools. The student samples were drawn using a computer-based system carried out by field staff and specified through the use of session assignment forms.

Field supervisors carried out the sampling of students a week before the assessment. Student listing forms were prepared in each participating school. All enrolled 12^{th}-grade students were to be entered on the form in any order convenient to the school. Before carrying out the sampling, a field supervisor reviewed the form and made comparisons with other enrollment information to ensure that the list included all eligible students. Once the list was determined to be complete, a sequential line number was assigned to each student.

The within-school student sample size varied by public/nonpublic school and enrollment size. For public schools, if the number of eligible 12^{th}-grade students on the student listing form was 110 or less, all students were selected. If the school had more than 110 12^{th}-grade students on the form, 100 students were selected. Because nonpublic schools, which generally have small enrollment, were heavily oversampled, they were allowed to be selected or "hit" more than once (actually up to three times) to limit the number of schools in the sample. The more 'hits' the school has, the more students are selected to be assessed. The grade-specific enrollment size of the school determines its number of 'hits,' and the number

of 'hits' determines the number of students within a school to select. If the number of students on the student listing form for nonpublic schools was less than or equal to 62 students per "hit," all students were selected. For those nonpublic schools with more than 62 students per "hit," 60 students per "hit" were selected. For example, if a school was selected or "hit" three times, and if the number of students in Grade 12 was less than or equal to 186 (62 * 3), then all students would be selected. If the school had 500 students, then only 180 students (60 * 3) would be selected. Guidelines were in place for both public and nonpublic schools to alleviate sampling burdens.

Some schools that were originally assigned with two sessions were found to have significantly fewer students than was expected at the time of sampling. In these cases one of the session types was randomly dropped proportional to the session type allocation. That is, in such public schools the probability of dropping the mathematics and science session was 7/16 and 9/16, respectively. In such nonpublic schools the probabilities were 50-50 for mathematics and science. The enrollment size cutoff for dropping a session was 28 for public schools and 24 for nonpublic schools.

The students selected in the initial sample were allocated to session types based on the number of sessions assigned to the school and whether the school was public or nonpublic. If a school was assigned only one session, all students were allocated to the session type assigned to the school. For public schools assigned with two sessions, 7 out of 16 students were systematically assigned to mathematics and the other 9 students were assigned to science. For nonpublic schools assigned with two sessions, every other student was assigned to mathematics and the remaining half was assigned to science.

In public schools with low minority enrollment, an oversample of Black and Hispanic students was selected. (The race/ethnicity of students was determined from school administrative records.) After the initial sample was selected, the Black and Hispanic students not selected were identified and listed. They were then sampled to a total that, in expectation, was the same number of Black and Hispanic students as were already selected. In practice, if the number of students not selected was less than the number of selected students, then all Black and Hispanic students not selected were to be assessed also. Otherwise, Black and Hispanic students were sampled so that their overall within-school probability of selection was twice the rate of other students. Since nonpublic schools are generally small and homogeneous, no oversampling of minority students was conducted for this study.

An additional oversample of SD/LEP students was selected for all schools. The general intent of this oversampling was to select SD/LEP students at twice the rate at which non-SD/LEP students

were sampled (or to include all SD/LEP students if there were not sufficient numbers to permit sampling at twice the rate). In each school, the initial sample of students was drawn from the full list of eligible students. In public schools in low-minority areas (i.e. less than 15 percent Black and Hispanic), an oversampling of Black and Hispanic students then occurred. Among those students in the school not selected for either of the two prior samples, the SD/LEP students were identified. A sample from among the identified SD/LEP students was drawn, using a sampling rate that would achieve the double sampling rate required overall.

The students selected in the SD/LEP and/or Black and Hispanic oversample procedure were also allocated to session types in the same fashion as the initial sample described above.

As part of the computer-based sampling system, a session assignment form was generated for each school where sampling was carried out and specified the students selected for sample. The form contained the following information:

- Number of students selected in the initial sample;
- Types of sessions that were to be administered at the school;
- Whether the school was eligible for Black and Hispanic oversampling;
- Line numbers (from the student listing form) specifying the students selected in the initial sample organized by session type;
- Line numbers for students selected for the Black and Hispanic oversample and/or the SD/LEP oversample organized by session type; and
- Special instructions as appropriate for the 2000 SD/LEP Questionnaire.

2.6 Students Not Included in the Assessment

Once the sample of students was selected, school staff members were asked to identify any students with a disability and any students classified as limited English proficient. The SD/LEP 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 members were also asked to determine whether any of the students identified as disabled or with limited English language proficiency 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 PRIMARY SAMPLING UNITS, SCHOOLS, AND STUDENTS FOR THE 2000 HIGH SCHOOL TRANSCRIPT STUDY

This chapter presents the sampling procedure used for the NAEP 2000 High School Transcript Study (HSTS 2000). Included are details describing the primary sampling units, the school sample, and the student sample.

The purposes of the NAEP 2000 High School Transcript Study were to gather data on a nationally representative sample of students who graduated from U.S. public and nonpublic high schools in 2000 and link the data to the NAEP 2000 national main assessment. For the HSTS 2000 sample of students to be as representative as possible, it included all public schools and a subsample of nonpublic schools found in the 12th-grade NAEP 2000 sampling frame. A representative sample of students was included from each school. When possible, the students selected for the transcript study were the same as those students selected for NAEP 2000. When this was not possible, a systematic sample of students was drawn from the school.

3.1 Primary Sampling Unit (PSU) Sample

All 94 PSUs selected for NAEP 2000 were retained for the HSTS 2000. This retention was a departure from previous HSTS studies, where only a subsample of the NAEP PSUs was used in order to reduce field costs.

3.2 School Sample

The HSTS 2000 school sample comprised all 319 12th-grade public schools and a subsample of the 621 12th-grade nonpublic schools selected for NAEP 2000. The objective of nonpublic school subsampling was to reverse the oversampling of nonpublic schools in NAEP 2000 so that the nonpublic school students in the HSTS 2000 were represented in proportion to their prevalence in the general 12th-grade student population. While an oversample of nonpublic schools was necessary for the NAEP 2000 sample to meet student sample requirements, it was not desirable for the HSTS 2000 sample. Nonpublic schools tend to be smaller than public schools, so collection cost per transcript is much higher. To reverse

the oversampling, nonpublic schools were subsampled differentially by reporting group with probability proportional to size (PPS), as shown in table 5.

Table 5. HSTS subsampling rates for nonpublic schools by reporting group: 2000

Reporting group	Subsampling rate (percent)	Number of schools selected for NAEP	Number of schools selected for HSTS
Total	†	621	60
Catholic	11.0	127	14
Lutheran	1.2	54	1
Conservative Christian	10.6	130	14
Other religious	10.7	110	12
Nonsectarian	6.5	123	9
Independent	13.1	52	7
Unknown affiliation	10.7	25	3

† Not applicable.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

3.3 Student Sample

For schools participating in both NAEP 2000 and HSTS 2000, the same students were included in the two samples where possible. For privacy reasons, the only means of identifying the students participating in NAEP 2000 was a list left in the school office. Since the NAEP assessments were administered from January through March 2000, the schools were asked to retain the NAEP Administration Schedules until the HSTS data collection in the spring and summer of 2000.[7] The Administration Schedules were forms produced specifically for each school. They included the assessment booklet IDs that were assigned to each school, which were listed sequentially on the forms. Once the student sample was drawn, the selected student's name was recorded on the Administration Schedule for the type of session for which he or she was selected. As this was done, the booklet ID on that line became the student's NAEP ID number. This form was the only place where selected students' names were recorded. To maintain the students' confidentiality, the part of the Administration Schedule with the students' names was never removed from the school. Other demographic information was also recorded on the Administration Schedule, which is shown in appendix A.

[7] NAEP asked schools to retain the administration schedules until the end of the calendar 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, 2000.

For schools that participated in NAEP 2000 but were missing their Administration Schedules, and for schools that agreed to provide transcripts but did not participate in the NAEP 2000 assessment, the field workers sampled the students using the following rules:

- If 60 or fewer students were in the senior class, all students were selected for the study.
- If more than 60 students were in the senior class, the field worker drew a systematic random sample of 50 students.

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 (50) 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 supplied 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 (TRF) (exhibit A-3 in appendix A) and gave it to the school staff to draw the transcripts. The TRF also provided a place to record the students' graduation status, gender, race/ethnicity, birth month, birth year, disability status, limited English proficiency (LEP) status, receipt of Title I services, and National School Lunch Program participation.

When field workers went to the schools to collect the transcript data, they took sets of labels for each student NAEP ID at the school. As they collected the transcripts, they attached the ID labels to them to identify the student to whom the transcript belonged. To maintain confidentiality, the field worker removed the students' names from the TRF before taking the form from the school along with the transcripts. They also made sure that any identifying information on the transcripts was either erased or obscured, so that the student could not be identified.

For schools that had not participated in NAEP 2000, a set of labels was created with newly assigned ID numbers for the students selected in that school. In those schools, the TRF was produced with the new ID numbers, but with space to record all of the demographic information that was collected.

A total of 23,440 students were selected for the HSTS 2000. Of these students, 22,010 students were from schools that maintained their NAEP 2000 administration schedules and were identified by their NAEP booklet numbers. Another 630 students were from schools that participated in

NAEP 2000 but had lost the link between student names and NAEP booklet numbers, and 800 students were from schools that did not participate in NAEP 2000.

Table 6 displays the number of eligible schools in the sample and the number and percentage of schools from which transcripts were collected, by linking category. Where it is indicated that transcripts were collected, it means they were usable transcripts of graduating students.

Table 6. Response rate of eligible schools by linking category, unweighted: 2000

School participation status	Number of schools in sample[1]	Number of schools where transcript data were collected	Percentage of schools where transcript data were collected
Total eligible schools in sample	**359**	**277**	**80.8**
Eligible original sampled schools	343	265	77.3
Original school participated in NAEP—IDs linked to NAEP IDs	258	236	91.5
Original school participated in NAEP—IDs not linked to NAEP IDs	13	13	100.0
Original school did not participate in NAEP	72	16	22.2
Eligible substitute schools	16	12	75.0
Substitute school participated in NAEP—IDs linked to NAEP IDs	16	12	75.0
Substitute school participated in NAEP—IDs not linked to NAEP IDs	0	0	0.0

[1]A sampled school was defined as the original school. When a substitute school replaced an original school, this replacement did not change the number of schools in the sample. The 20 ineligible schools in the sample were not included in this table.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

Because sampling was performed in most high schools prior to graduation, not all sampled students were, in fact, graduates. Only graduates, however, were eligible for the transcript study. From the exit status of the students, it was determined that of the 23,440 students in the sample, 21,085 actually graduated by October 2000 and 2,355 did not. From the 21,085 graduates, 20,931 transcripts were collected and processed. That is, 99.3 percent of the transcripts of eligible students were obtained. Table 7 displays the number of sampled students in the participating (original and substitute) schools and the number and percentage of completed transcripts of graduates that were processed.

Table 7. Percentage of sampled students who were graduates and for whom completed transcripts were received: 2000

School participation status	Number of students in sample	Number and percentage of sampled students who were graduates and for whom completed transcripts were received[1]	
		Number	Percentage
All schools	23,440	20,931	89.3
School participated in NAEP— IDs linked to NAEP IDs	22,010	19,547	88.8
School participated in NAEP— IDs not linked to NAEP IDs	630	609	96.7
School did not participate in NAEP	800	775	96.9

[1]This number reflects the number of usable transcripts collected.
SOURCE: U.S Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

Table 8 displays the unweighted response rates for graduates in the eligible participating schools. Table 9 displays the weighted response rates for NAEP, the transcript study, and the linked schools.

Table 8. Response rates of graduates, unweighted: 2000

School participation status	Known graduates	Number of transcripts of known graduates collected	Percentage of transcripts of known graduates collected
All schools	21,085	20,931	99.3
School participated in NAEP— IDs linked to NAEP IDs	19,691	19,547	99.3
School participated in NAEP— IDs not linked to NAEP IDs	612	609	99.5
School did not participate in NAEP	782	775	99.1

SOURCE: U.S Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

Table 9. Response rates for NAEP, transcript study, and linked schools, weighted: 2000

	Weighted school response rate before substitution (percent)	Weighted school response rate after substitution (percent)	Weighted student response rate (percent)	Overall response rate (percent)
Overall NAEP				
Mathematics R2	78.3	82.4	76.6	63.2
Mathematics R3	78.3	82.4	77.2	63.6
Science R2	77.4	81.9	75.9	62.2
Science R3	77.4	81.9	75.9	62.2
Transcript Study				
Overall	78.5	81.9	99.4	81.5
NAEP participating schools	89.0	93.3	99.4	92.7
Linked Schools				
Mathematics R2	78.4	81.7	79.9	65.2
Mathematics R3	78.4	81.7	80.1	65.4
Science R2	78.4	82.1	79.2	64.0
Science R3	78.4	82.1	78.9	64.7

NOTE: The R2 reporting sample is the nonaccommodated reporting sample. Sampled students include students who have neither a student disability (SD) nor a limited English proficiency (LEP), plus SD/LEP students from sessions in which accommodations were not allowed. The R3 reporting sample is the accommodated reporting sample. Sampled students include students who have neither a student disability nor a limited English proficiency, plus SD/LEP students from sessions in which accommodations were allowed.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

For the NAEP-participating schools in the 2000 transcript study, the weighted school response rate equaled 93.3 percent, while their weighted student response rate equaled 99.4 percent. The overall response rate for the 2000 transcript study's NAEP students equaled 92.7 percent. When factoring in NAEP school nonresponse into the 2000 transcript study, the weighted school response rate equaled 81.9 percent, while the weighted student response rate equaled 99.4 percent. The overall response rate for the 2000 transcript study equaled 81.5 percent.

The HSTS 2000 attained both school and student response rates that were below 85 percent. According to NCES standards, any survey not achieving an 85 percent response rate must provide a nonresponse bias analysis. As the NAEP 2000 12th-grade assessment, of which HSTS 2000 is a component, also had school and student response rates below 85 percent, and because the HSTS 2000 response rates for NAEP-participating schools was above 85 percent, the nonresponse bias analysis for the NAEP 2000 12th-grade assessment would satisfy this requirement.

Following completion of the weighting for the NAEP assessments, a 2000 nonresponse bias study was conducted to determine if the 2000 data may have been biased by differential nonresponse that might explain, in part, the decline in science scores between 1996 and 2000. This report includes tables from this nonresponse bias analysis for the 12th-grade science assessment. For purposes of comparison with previous NAEP studies, tables 10 and 11 compare the NAEP nonresponse rates after school substitution from the 1996 and 2000 studies. Tables 12 and 13 compare the 2000 responding and nonresponding schools by school and student demographics. It was determined that the effects of school and student nonresponse were not sufficient as to result in suppression or annotation of the NAEP 12th-grade results.

Table 10. Weighted after substitution school response rates, national main NAEP grade 12 science samples: 1996 and 2000

	1996			2000		
Population	Sample size	Response rate	Standard error	Sample size	Response rate	Standard error
NAEP Region						
Northeast	62	76.7	6.5	164	77.0	5.0
Southeast	77	60.5	9.2	160	85.1	3.9
Central	67	74.5	7.9	154	87.4	2.7
West	87	80.7	6.0	212	81.9	4.5
School type						
Catholic	32	87.5	6.0	125	92.5	3.1
Other nonpublic	38	47.7	11.7	278	63.0	4.9
Public	223	79.3	3.7	287	88.7	2.1
School size						
1–49	44	62.8	9.4	214	82.3	3.5
50–399	195	77.0	4.2	389	84.5	2.4
400+	54	86.3	5.5	87	83.0	4.5
School location						
Large city	52	91.4	3.8	188	82.3	4.2
Midsize city	47	65.8	10.8	91	82.9	5.9
Urban fringe/large city	49	62.4	7.2	239	75.3	4.3
Urban fringe/midsize city	28	73.6	7.5	56	75.0	6.4
Large town	3	15.5	16.9	3	81.1	82.1
Small town	60	76.1	8.9	50	70.9	7.1
Rural	54	73.1	7.5	63	95.7	2.0
Minority status						
High Black/Hispanic public	126	87.2	3.1	182	88.3	2.7
Low Black/Hispanic public	97	75.9	4.9	105	88.8	2.8
Nonpublic	70	56.4	9.5	403	69.2	4.1

NOTE: The weighted rates use school base weights alone, unlike NAEP's traditional school-level response rates, which incorporate student enrollment as well.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

Table 11. Weighted after substitution student response rates, national main NAEP grade 12 science samples: 1996 and 2000

Population	1996 S2 reporting population			2000 R2 reporting population		
	Sample size	Response Rate	Standard error	Sample size	Response Rate	Standard error
Overall	**9,806**	**77.5**	**1.7**	**18,985**	**75.9**	**1.3**
NAEP region						
Northeast	2,075	77.6	3.2	4,546	72.5	2.4
Southeast	2,552	83.5	3.8	4,728	78.7	2.3
Central	2,179	75.2	3.7	3,965	78.4	2.5
West	3,000	74.6	2.9	5,746	74.3	2.3
School type						
Catholic	1,017	91.4	2.3	3,242	87.5	1.4
Other Nonpublic	557	90.6	2.1	4,096	90.5	1.2
Public	8,232	76.0	1.8	11,647	74.7	1.4
School location						
Large city	2,105	74.0	4.3	5,126	70.8	2.8
Midsize city	1,828	69.0	4.0	2,417	67.4	3.8
Urban fringe/large city	1,565	72.6	4.3	6,747	72.3	2.7
Urban fringe/midsize city	1,013	73.3	4.3	1,593	79.2	3.0
Large town	57	75.4	‡	105	68.2	‡
Small town	1,914	83.9	2.8	1,530	87.6	2.6
Rural	1,324	85.5	3.6	1,467	86.7	2.2
Age category						
At modal age or younger	7,827	78.5	1.8	15,557	76.2	1.3
Older than modal age	1,979	73.3	1.9	3,428	74.8	1.7
Race/ethnicity category						
White	6,055	77.7	2.0	12,128	76.2	1.5
Black	1,644	76.2	3.2	2,831	72.1	2.5
Hispanic	1,432	74.3	2.5	2,821	76.9	2.4
Other	675	81.8	2.9	1,205	78.8	2.5
Gender						
Missing	2	‡	‡	105	15.6	9.2
Male	4,697	76.5	2.0	9,100	75.9	1.3
Female	5,107	78.5	1.6	9,780	76.2	1.3
SD						
Yes	296	67.4	3.4	566	69.7	3.6
No	9,510	77.8	1.7	18,419	76.2	1.3
LEP						
Yes	255	72.0	3.1	218	78.7	5.7
No	9,551	77.6	1.8	18,767	75.9	1.3
SD, LEP						
SD yes, LEP yes	10	37.8	13.7	14	83.7	7.4
SD yes, LEP no	286	68.0	3.5	552	69.4	3.6
SD no, LEP yes	245	73.3	3.0	204	78.4	5.9
SD no, LEP no	9,265	77.9	1.8	18,215	76.2	1.3

‡ Could not be computed due to insufficient sample size.

NOTE: The weighted response rates use student base weights, which do not include an adjustment for school nonresponse. The 1996 S2 and 2000 R2 reporting populations both define the nonaccommodated reporting population. The population includes students who have neither a student disability (SD) nor a limited English proficiency (LEP), plus SD/LEP students from sessions in which accommodations were not allowed.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

Table 12. Weighted distributions (in percents) of responding and nonresponding schools for the national main NAEP 2000 sample for grade 12 science

	All schools			Responding schools			Nonresponding schools		
Population	Sample size	Percent	Standard error	Sample size	Percent	Standard error	Sample size	Percent	Standard error
NAEP Region									
Northeast	164	18.9	1.7	130	17.5	1.8	34	26.2	5.5
Southeast	160	20.8	2.2	131	21.2	2.5	29	18.7	4.6
Central	154	32.7	2.4	121	34.3	2.9	33	24.9	4.6
West	212	27.5	2.5	168	27.0	3.0	44	30.1	5.8
School type									
Catholic	125	5.6	0.7	115	6.2	0.8	10	2.6	1.1
Other nonpublic	278	21.3	2.0	192	16.1	2.1	86	47.5	5.4
Public	287	73.1	2.2	243	77.7	2.3	44	49.9	5.6
School size									
1–49	214	44.0	3.3	158	43.4	3.8	56	47.0	5.7
50–399	389	48.6	3.2	320	49.3	3.7	69	45.4	5.5
400+	87	7.4	0.9	72	7.4	1.0	15	7.6	2.4
School location									
Large city	188	11.6	1.4	151	11.4	1.5	37	12.3	3.2
Midsize city	91	9.4	1.7	74	9.3	1.9	17	9.7	3.2
Urban fringe/large city	239	22.2	2.4	191	20.0	2.4	48	33.0	5.4
Urban fringe/midsize city	56	8.0	1.6	43	7.2	1.8	13	12.1	3.2
Large town	3	0.6	0.5	2	0.6	0.6	1	0.7	0.7
Small town	50	13.2	1.4	34	11.2	1.4	16	23.1	5.9
Rural	63	35.0	3.2	55	40.2	3.6	8	9.0	4.0
Minority status									
High Black/Hispanic public	182	22.2	2.1	156	23.5	2.6	26	15.7	3.1
Low Black/Hispanic public	105	50.9	3.3	87	54.2	3.8	18	34.3	5.9
Nonpublic	403	26.9	2.2	307	22.3	2.3	96	50.1	5.6

NOTE: The weighted distributions represent school base weights alone. Details may not sum up to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

Table 13. Weighted distributions (in percents) of eligible responding and nonresponding students for the national main NAEP 2000 sample for grade 12 science

	All students			Responding students			Nonresponding students		
Population	Sample size	Percent	Standard error	Sample size	Percent	Standard error	Sample size	Percent	Standard error
NAEP region									
Northeast	4,546	22.3	1.8	3,562	21.3	1.7	984	25.5	3.2
Southeast	4,728	25.0	1.6	3,910	26.0	1.7	818	22.1	2.6
Central	3,965	22.8	1.6	3,205	23.5	1.8	760	20.4	2.5
West	5,746	29.9	2.0	4,432	29.3	2.2	1,314	32.0	2.9
School type									
Catholic	3,242	5.7	0.5	2,848	6.6	0.6	394	3.0	0.5
Other Nonpublic	4,096	3.2	0.2	3,699	3.8	0.3	397	1.2	0.2
Public	11,647	91.1	0.6	8,562	89.7	0.7	3,085	95.8	0.5
School location									
Large city	5,126	17.3	1.8	3,908	16.2	2.0	1,218	21.0	2.4
Midsize city	2,417	11.2	1.7	1,876	10.0	1.5	541	15.2	3.2
Urban fringe/large city	6,747	33.7	2.7	5,320	32.1	2.8	1,427	38.8	3.9
Urban fringe/midsize city	1,593	11.5	2.5	1,306	12.0	2.6	287	10.0	2.7
Large town	105	1.2	1.2	72	1.1	1.1	33	1.6	1.6
Small town	1,530	11.0	1.7	1,347	12.7	2.0	183	5.7	1.6
Rural	1,467	14.0	2.1	1,280	15.9	2.3	187	7.7	2.0
Age category									
At modal age or younger	15,557	80.4	0.7	12,455	80.7	0.8	3,102	79.5	1.1
Older than modal age	3,428	19.6	0.7	2,654	19.3	0.8	774	20.5	1.1
Race/ethnicity category									
White	12,128	66.4	1.8	9,763	66.7	1.8	2,365	65.7	2.6
Black	2,831	13.3	1.2	2,115	12.7	1.2	716	15.5	1.9
Hispanic	2,821	13.5	1.3	2,242	13.6	1.4	579	12.9	1.6
Other	1,205	6.8	0.6	989	7.0	0.6	216	5.9	0.8
Gender									
Missing	105	0.2	0.1	47	0.0	0.0	58	0.5	0.3
Male	9,100	48.8	0.5	7,242	48.8	0.6	1,858	48.9	0.9
Female	9,780	51.0	0.5	7,820	51.2	0.6	1,960	50.5	0.9
SD									
Yes	566	4.4	0.5	394	4.1	0.5	172	5.6	0.8
No	18,419	95.6	0.5	14,715	95.9	0.5	3,704	94.4	0.8
LEP									
Yes	218	1.5	0.4	171	1.5	0.4	47	1.3	0.5
No	18,767	98.5	0.4	14,938	98.5	0.4	3,829	98.7	0.5
SD, LEP									
SD yes, LEP yes	14	0.1	0.0	11	0.1	0.1	3	0.1	0.0
SD yes, LEP no	552	4.3	0.5	383	3.9	0.5	169	5.5	0.8
SD no, LEP yes	204	1.4	0.4	160	1.4	0.3	44	1.2	0.5
SD no, LEP no	18,215	94.2	0.6	14,555	94.5	0.6	3,660	93.2	0.9

NOTE: The weighted response rates use student base weights, which do not include an adjustment for school nonresponse. Details may not sum up to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

4. DATA COLLECTION PROCEDURES

This chapter discusses the procedures used in the data collection for the 2000 High School Transcript Study. Included are sections on field worker training, contacts with schools, and obtaining course catalogs and transcripts.

4.1 Training NAEP 2000 Field Supervisors as Data Collectors

The field workers for the 2000 High School Transcript Study (HSTS 2000) were drawn from the pool of 2000 National Assessment of Educational Progress (NAEP 2000) field supervisors. They were trained in the data collection procedures for HSTS 2000 in December 1999. Conducted by the HSTS 2000 curriculum specialist/coding supervisor, the training consisted of three sessions which took a full day to complete.

The purpose of the first session was to establish the background knowledge needed to help field workers 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 2000 data. The first training session consisted of a presentation describing the purposes of the HSTS 2000, the procedures to be used in handling and processing HSTS 2000 data, and the most appropriate school sources to use in obtaining needed data. Specific examples were used throughout the presentation.

The second training session was held to familiarize field workers with the HSTS 2000 materials and forms and with the variety of materials they could expect to find in the schools. During the second session, field workers were shown examples of various types of high school records and materials, including school- and district-level catalogs, course lists, transcripts, and all the forms used for the HSTS 2000. The field workers learned how the information on each of these materials became the data needed at the school and student levels. Transparencies of screen prints of the transcript data entry and course coding systems were shown to them to demonstrate how the information from the specific material would be entered into the systems by data entry staff.

The third session provided an opportunity for field workers to work with sample catalogs and transcripts, and to fill out practice forms similar to actual materials used for the HSTS 2000. The third training session consisted of completing sets of exercises, designed 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 HSTS 2000, with all identifying information deleted.

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 pairs or 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 before the training session ended. 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, the physical layout of the materials, and the ease or difficulty of accessing the information in the materials. Transcripts were examined to show a number of ways that the following courses might be listed or described:

- special education courses;
- transfer courses;
- remedial courses;
- honors courses;
- off-campus location courses; and
- courses for students with limited English proficiency.

4.2 Contacts with States, Districts, and Schools

In September 1999, superintendents and principals were notified about the transcript study through the Summary of School Activities (see exhibit A-7 in appendix A), which was included in a mailout to all schools selected for NAEP 2000. The summary provided information about participating in the HSTS 2000, including procedures that would be used to ensure confidentiality of the data, and the amount and nature of school staff time required for participating in HSTS.

In December 1999, district superintendents of participating public 12th-grade schools sampled for the main NAEP and selected for the HSTS 2000 were mailed additional information concerning the HSTS. Items in the package included the following:

- An informational letter to school superintendents from NCES (see exhibit A-8 in appendix A)
- A list of schools in the district selected for the HSTS 2000; and
- A Summary of School Activities.

Once participation in the study was authorized by the district, the individual public schools were contacted. Private schools were contacted directly since no higher level authorization was required.

For contacts with both public and private school personnel, field workers followed the same procedures. They were provided with the following materials:

- An informational letter to principals from NCES (see exhibit A-9 in appendix A); and
- A Summary of School Activities.

Field workers provided these materials to the school principals and school coordinators during their initial visit to schools. They discussed the HSTS 2000 with the school coordinator prior to the sampling visit when they called to confirm the sampling date.

Initial HSTS information requested from schools included information school personnel were asked to provide on the School Information Form (SIF), as well as their school's course catalogs for the four most recent school years, including 1999–2000, and three sample transcripts. This initial information was collected by field workers at the time of their first visit. The schools were also asked to provide a complete transcript for each graduate in the HSTS 2000 sample as soon as graduation information was posted on the transcripts. Information provided on the SIF indicated the appropriate date for the HSTS 2000 field workers to obtain the transcripts.

For eligible participating NAEP schools that agreed to cooperate, students sampled for NAEP 2000 were included in the HSTS 2000 sample, and a brightly-colored Disclosure Notice was placed in their folder by a NAEP 2000 field worker or school staff member. This notice served two functions:

- It alerted the school personnel that information contained in the student's folder would be used for the HSTS 2000.
- Because of its color, it also served as a visible marker for identifying the folders of students in the HSTS 2000 sample to facilitate finding their transcripts later.

Notification to the originally nonparticipating NAEP 2000 schools included information that the intent was to select a sample of up to 50 students and to provide the same confidentiality safeguards with these samples as with all NAEP students. That is, student names would be removed from any papers that left the school. Field workers also emphasized that a school's participation in the HSTS 2000 would not involve any student time.

For both NAEP 2000 participating and nonparticipating schools, the initial contact by the field worker included a discussion of the following:

- Procedures for obtaining transcripts for the selected students and the method for reimbursing the school for the expense; and
- The availability of a course catalog or description.

An appointment was then set to visit the school to prepare the transcript requests and obtain the course catalogs.

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

Field workers requested sample materials for the HSTS 2000 when they first contacted a school and collected these materials when they visited the school for sampling. There were 264 schools that participated in both NAEP 2000 and HSTS 2000 (although 13 of these schools did not maintain the NAEP-HSTS links). There were also 16 schools from the original school sample that participated in the HSTS 2000, but did not participate in NAEP 2000. The sample materials included, preferably, a course catalog (a list of courses) offered for each of four consecutive years, from 1996–1997 through 1999–2000; a completed School Information Form; and three sample transcripts, one representing a "regular" student, one with honors courses, and one with special education courses. Since these materials were unique to each school, acquiring them before the collection of the actual transcripts enabled HSTS 2000 staff to examine them and call a field worker or the school (e.g., before school personnel left for the summer) with any questions that arose during the school year.

The field worker also gathered general information about class periods, course credits, graduation requirements, and other aspects of school policy. Sometimes this information was documented in the course catalog and at other times in a separate school policy document.

4.3.1 Catalogs

Course catalogs were carefully reviewed at the school. Field workers verified that the catalogs contained all of the courses that 12th-graders could have taken in high school, including vocational, remedial, honors, special education, or off-campus courses, or courses taught in a language other than English. If these course listings were not in the catalog, every effort was made to obtain additional information from school personnel to document the existence of such courses and to describe them.

The HSTS requests course catalogs that contain the most comprehensive information about the courses offered by the schools. Ordered from most to least complete, the requested types of catalogs are as follows:

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;
4. A school-level course list without descriptions; or
5. A district-level catalog without any indication of which courses were offered in specific schools.

All catalogs and course lists that were received by field workers were forwarded to HSTS 2000 data processing staff.

4.3.2 Sample Transcripts

Since transcript format varied greatly among school districts throughout the country, it was sometimes difficult to find the needed information on a transcript. This difficulty 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, three transcripts of previous graduates were obtained from each school by the NAEP field workers during the NAEP 2000 assessment. The three transcripts requested from each school included one that contained honors-level courses, one that contained special education courses, and one that contained just the "regular" courses. The HSTS field workers marked each transcript to indicate where on the transcript the needed information was found and how information regarding course level was coded. Attached to each marked-up transcript was a Transcript Format Checklist (exhibit A-10 in appendix A) indicating the key transcript information and whether or not that information was found or found and marked on the school's transcripts.

4.3.3 School Information Form

The School Information Form (SIF) was forwarded for data processing along with the other preliminary materials as described above. The SIF was completed by the field worker or a school staff member or sometimes by both. The name and position of the school's HSTS 2000 coordinator who helped fill out the SIF appeared on the first page. Along with general school information, the completed SIF contained the following information:

- sources of information within the school (if needed to complete HSTS 2000 data collection);
- the course description materials;
- graduation requirements;
- grading practices at the school; and
- 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 Questionnaire

The School Questionnaire (see appendix B) is a NAEP 2000 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 2000 for the NAEP participating schools. Schools that did not participate in NAEP 2000 were given a School Questionnaire to complete by field workers during the data collection phase of the HSTS 2000.

4.4 Identifying the Sample Students and Obtaining Transcripts

The HSTS 2000 used the NAEP 2000 sample for selecting schools and students in NAEP participating schools. For schools that participated in NAEP 2000, the student sample was recorded on the NAEP 2000 Administration Schedules. For schools that did not participate in NAEP 2000, the field worker drew a sample of students at the school. Details on how this sample was drawn can be found in section 3.3. The procedures for identifying students in schools with NAEP 2000 materials and in schools without NAEP 2000 materials are described in detail in separate sections that follow.

4.4.1 Schools with NAEP 2000 Materials

Schools that participated in NAEP 2000 identified students participating in the HSTS 2000 at the same time that the NAEP 2000 sample was selected. For all HSTS 2000 participants, a brightly colored Disclosure Notice 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.

Transcripts were requested for all students who were sampled for NAEP 2000. They included all assessed students, sampled students who were absent during the NAEP assessment, and students with disabilities (SD) and limited English proficiency (LEP) students who were excluded by the school from participating in the assessment.

When graduation information was posted on transcripts, a field worker returned to the school to obtain the requested transcripts. That date was provided by the school on the School Information Form. For each NAEP 2000 school, the field worker was given a Transcript Request Form (TRF) (see exhibit A-3 in appendix A). In addition to student name and NAEP ID, it contained columns for entering graduation status, gender, birth month and year, race/ethnicity, SD status, LEP status, Title 1 participation, and National School Lunch Program participation. Data available from NAEP 2000 files (NAEP ID and demographic variables) were preprinted on the form. The completed TRFs contained the following information:

- **Student Name –** The field worker recorded the first name, middle initial, and last name of each assessed, absent, or excluded student listed on the NAEP 2000 Administration Schedule. These entries were made to correspond to the preprinted NAEP ID.

- **NAEP ID –** The 10-digit NAEP 2000 assessment booklet numbers and SD/LEP questionnaire numbers for students excluded from the assessment were preprinted in ID order. This column on the TRF identified all students for whom transcripts were needed.

- **Exit Status –** Using information provided by the school, field workers assigned one of the following codes to describe each student's outcome at the school:

  1. Graduated with a standard diploma;
  2. Graduated with an honors diploma;
  3. Received a diploma with special education adjustments;
  4. Received a certificate of attendance;
  5. Still enrolled in this school;
  6. Dropped out;
  7. Other, such as transferred, Graduate Equivalency Diploma, or unknown;
  8. Out of scope; or
  9. Completed course requirements but did not pass required graduation tests.

  Sometimes the exit status was determined directly from the transcripts, and sometimes it was determined by other records or provided by school personnel.

- **Birthdate, Gender, and Race/Ethnicity –** Demographic information was generally preprinted for each sampled student. If not preprinted, it was recorded from the NAEP 2000 Administration Schedule. If the school informed a field worker that some of this information was incorrect, the field worker entered the correct information on the TRF.

- **SD and LEP Status –** For each student, it was recorded whether or not the student was classified by the school as SD and/or LEP.

- **National School Lunch Program and Title 1 –** Field workers recorded yes or no for participation in each of these programs.

- **Transcript Received –** Field workers checked this column to document that the transcript for a given student had been received.

Once the TRF was completed by carefully transferring student information from the Administration Schedules, the field worker filled out the summary box at the top of the form and requested transcripts according to the procedures set forth by the school. As already noted, the Disclosure Notice placed in students' folders at the time of the NAEP 2000 assessment helped to facilitate transcript collection in participating NAEP schools.

Once the field worker filled in the names of the students, some schools were able to access an electronic data file and print the transcripts. In other schools, the school coordinators pulled transcripts from their folders and photocopied them at the school.

When the request was filled, the field worker reviewed the transcripts to ensure that a transcript was received for each 12^{th}-grade student selected for the NAEP 2000 assessment, whether or not that student had graduated. Even though nongraduate transcripts were not included in the HSTS, each student graduation status needed to be accounted for and verified. Each transcript was checked for eligibility, understandability (e.g., are all the codes on it defined on the transcript or explained in the SIF?), and completeness. The field worker then labeled each transcript with preprinted labels containing the School ID and the NAEP ID for the student. The field worker completed a Documentation of Missing Transcripts 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 the data processing staff. 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.

Personal identifiers were also removed from the Transcript Request Forms. Before sending the TRFs from the school, the field worker cut off the portion that contained the students' names to comply with confidentiality provisions. The portion with the names was left in the school's NAEP folder.

Schools were reimbursed at their standard rates for providing the transcripts.

4.4.2 Schools without NAEP 2000 Materials

In schools that did not participate in NAEP 2000, the field worker first selected a sample of students, then requested transcripts for those students and followed the procedures described in the previous section for reviewing and shipping transcripts. The School Information Form was also completed, and course catalogs for the past 4 academic years were collected. The school was also asked to complete the NAEP 2000 school questionnaire. The information in the catalogs was documented by completing the Course Catalog Checklist (exhibit A-11 in appendix A). At this point, the procedure was different. Rather than obtaining and annotating three sample transcripts, as was done at the time of the NAEP 2000 visit to the school, the field worker used the Transcript Format Checklist to annotate three actual transcripts from among those that were collected.

For the schools that participated in HSTS 2000 but not in NAEP 2000, 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 2000. This list was requested during the preliminary call placed to the school when it was determined that the school would participate in HSTS 2000. Information collected for each student selected to participate in HSTS 2000 included the information needed to complete the Transcript Request Form, as outlined in the section above (with exception of the NAEP ID). These data were collected either with the list of 2000 graduates or after sampling, depending on which procedure was easier for the school. The SD/LEP Questionnaires were not collected for students in schools that had not participated in NAEP 2000.

As described in section 3.3, there were two basic sampling rules for the HSTS 2000. These rules applied to all schools that required a new sample of students.

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

Because the students in the HSTS 2000-only schools did not have NAEP 2000 identification numbers, a set of IDs was preassigned for up to 60 students in each school. The field worker, with the assistance of the school, completed the Transcript Request Form (Version 2) and submitted it to the school staff. The transcripts then were provided to the field worker, who reviewed and shipped them to the data processing staff in the same manner as transcripts from schools participating in NAEP 2000.

4.5 SD/LEP Questionnaire

The questionnaire that NAEP 2000 uses to collect information from school staff about students with disabilities and students with limited English proficiency is called the SD/LEP Questionnaire (see appendix B). Schools were asked to have the person most knowledgeable about a disabled or limited English proficient student complete the 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. For schools participating in the NAEP 2000, the SD/LEP Questionnaires were collected as part of the NAEP procedures.

4.6 Sending Data for Processing

As with NAEP 2000, 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 2000 information was available, the ID number was the same as the student's NAEP 2000 booklet number. In schools where a sample of students was drawn specifically for the HSTS 2000, new IDs were generated.

After transcripts were collected and all information on sampled students recorded, field workers prepared the transcripts for transmittal to the data processing staff. They first compared the student ID and name 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 cut off the left hand column of the TRF, which contained the names of the students. The list of names remained in the schools (and was ultimately destroyed) and the remainder of the TRF was placed in the package to send to the HSTS 2000 field officer for data processing.

The field workers masked all personal 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, 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, and ZIP code), phone numbers, Social Security numbers, and other student ID numbers.

A Shipping Transmittal Form (exhibit A-12 in appendix A) accompanied all shipments to the data processing staff and summarized the types and number of materials being sent. This form also gave information on whether the transcripts were from the NAEP 2000 list or a new sample and, if the school did not participate in NAEP 2000, whether course catalogs and a School Information Form were included in the shipment.

4.7 Receipt and Review of Data from Data Collectors

When transcript study materials arrived for data processing, 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 two weeks of having their materials received for data processing.

An automated receipt system was developed and maintained by HSTS 2000 staff. 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.

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. Each school folder included the school's catalog or catalogs, Transcript Request Forms, student transcripts, Course Catalog Checklist, Transcript Format Checklists, School Information Form, and Shipping Transmittal Form.

Catalogs, sample transcripts, and School Information Forms were reviewed by the receipt clerk 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

The data from the 2000 High School Transcript Study (HSTS 2000) were processed through the student sampling information system, the Computer Assisted Data Entry System (CADE), and the Computer Assisted Coding and Editing System (CACE) simultaneously. To ensure the accuracy and consistency of data entry and coding, procedures were developed for the tasks described in detail in sections 5.1 through 5.9.

5.1 Establishing Student ID Control Lists

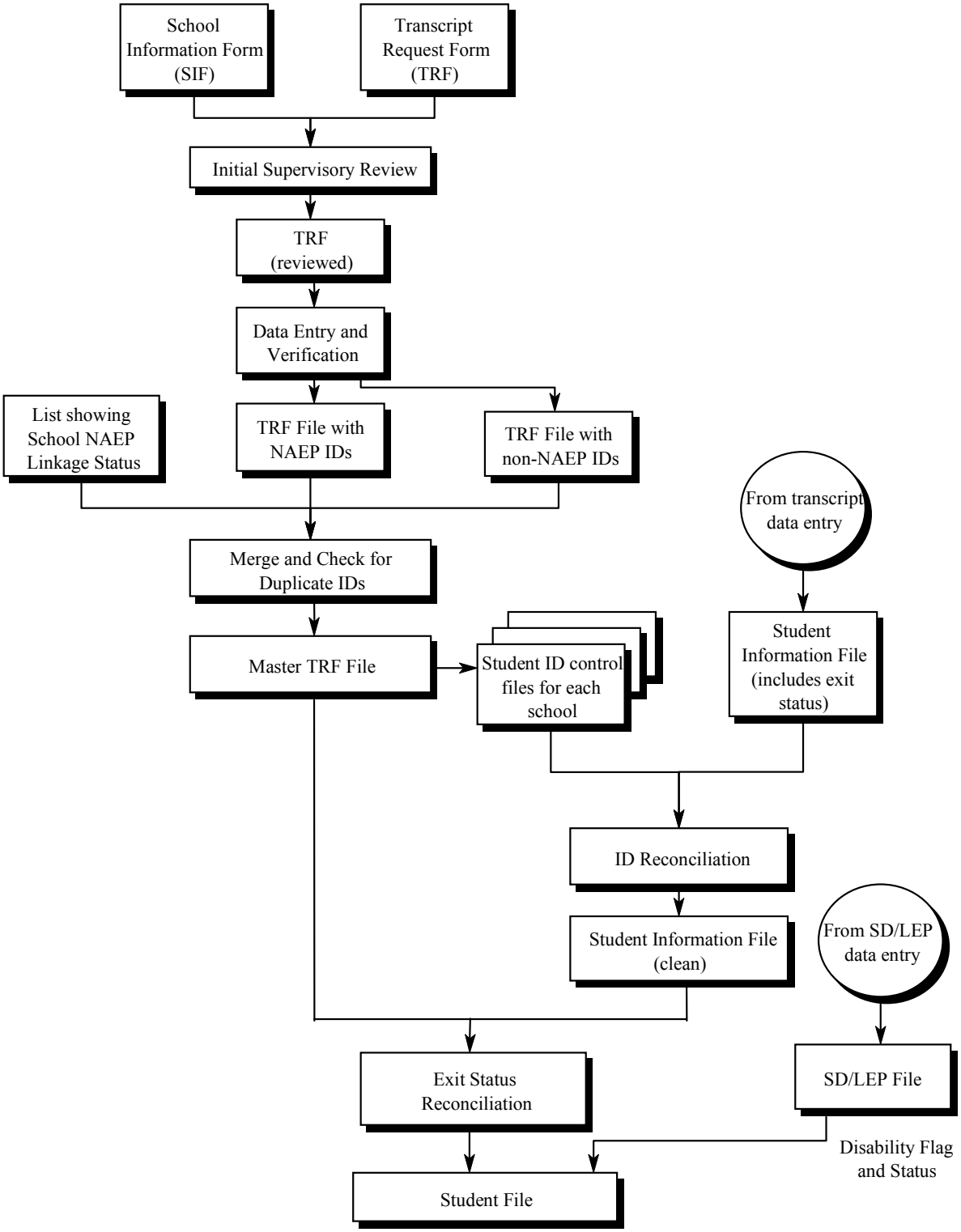
Student ID control lists were developed from lists obtained from the NAEP 2000 administration records for schools that participated in NAEP 2000. The control list for a school is the master list of IDs against which all other operations are checked. Only IDs matching those on the control lists are processed, 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 2000 students who were identified during the NAEP 2000 administration as 12^{th}-graders were retained on the control lists generated from NAEP 2000. Students identified as 10^{th}- or 11^{th}-graders, or those with an unknown grade, were removed from the lists.

For schools that did not participate in NAEP 2000, or that had lost the linkage between the students' names and their IDs, control lists were compiled from completed Transcript Request Forms (Version 2). A data file was created for each such HSTS 2000 school, listing the valid student IDs for that specific school.

5.1.1 Student Sampling Information System

The Transcript Request Form (TRF) and the sampling section of the School Information Form (SIF) provided the student sampling information for each school participating in the study. Figure 1 illustrates the process for entering the student sampling information. The figure also illustrates how intermediate files were used to ensure that all information was valid and that only valid student ID numbers were used.

Figure 1. Student information processing and ID reconciliation



5.1.2 School Information Form

In HSTS 2000 schools that also participated in NAEP 2000, the student sampling rates were identical to those used in NAEP 2000 because the sample was identical. For the 29 schools in which field staff drew samples in the field, the number of students listed (i.e., the number of eligible seniors) and the number of students sampled was recorded in the sampling section of the SIF. This information was keyed into a file that was checked against the number of unique student IDs on the TRF and then used in the weighting process.

5.1.3 Transcript Request Form

The preprinted information on the TRF was drawn from the NAEP 2000 student file. For schools that kept their NAEP 2000 materials, data entry was uncomplicated. The preliminary processing staff first created a file containing the preprinted information from the TRF with one record per student. Each student's graduation status as indicated on the TRF was entered at the end of each record. If necessary, the demographic data preprinted on the TRF was corrected. All entries were then key-verified; that is, re-keyed and matched up with the original keyed entry to catch and correct data entry errors. Finally, the staff key entered and key-verified all the TRFs from the schools for which new samples were drawn in the HSTS 2000 study.

The NAEP 2000 and non-NAEP 2000 TRF files were merged and checked for valid IDs and duplicates. Information in the TRF file and receipt control file was used to create a list of valid school identifiers with a flag indicating each school's linkage status to NAEP 2000. The linkage flag (LINKED in the restricted data school file) had four possible values:

0 = School did not participate in HSTS 2000;

1 = Both school ID and student IDs linked to NAEP 2000;

2 = School participated in HSTS 2000 only; and

3 = School participated in NAEP 2000 but, because a new sample was drawn, the student IDs did not match the NAEP 2000 booklet numbers.

The TRF file was also used to create a list of all valid student IDs within each school. These lists were key control mechanisms that were used throughout all phases of the study to ensure that only

valid IDs could be attached to each data record. For example, during entry of the transcript data, one of the data entry clerk's first steps was to key in the school ID and a student ID. As these IDs were keyed, the CADE system checked the IDs against the control lists and refused to accept any IDs not listed.

5.2 CADE System for Entering Transcript Data

The MS-Access-based Computer Assisted Data Entry (CADE) system included three basic levels of data entry, namely the school level, the student level, and the transcript level. The school-level data entry was handled by the School Materials and Information Component. It consisted of three screens: a School Receipt Control screen, a School Information screen, and a Transcript Receipt Control screen.

- The **School Receipt Control screen** recorded all material sent by the schools such as type of catalogs received, the number of transcripts requested, the number of transcripts received, and the types of diplomas or programs the school offered. The number of transcripts received from a school was matched to the number of transcripts that were processed throughout the different data entry and coding phases to reflect the progress of the different phases, provide accurate reports, and flag any outstanding or erroneous transcripts. A phase was completed only when all of a school's transcripts were processed. Verification could not commence until the data entry phase was completed, coding quality control checks could not be run until the verification phase was completed, and so on.

- The **School Information screen** recorded school-related information for standardization purposes. This information included the number of credits received for year-long courses (thus determining the Carnegie Conversion Factor or Carnegie Unit), the number of credits required for graduation in each subject area, a grade standardization scale, whether or not a state or district test was required for graduation, and which special programs were offered by each school. This information was used mainly in the data processing phase of the study.

- The **Transcript Receipt Control screen** recorded and tracked each student transcript that was received from a school and verified the student's exit status. The data entry staff identified whether or not a transcript was available to enter and made sure that the preloaded exit status reflected the correct graduation status of the student. Once this list of received transcripts was completed, it was used as a reference for data entry and verification completion.

The other two levels of data entry—student-level and transcript-level—was handled by the Student and Transcript Components of the system, each using a different dedicated screen. The Student Information screen recorded student-level information such as graduation date; rank in class; days absent each year; GPA as it appeared on the transcript; number of credits received, earned, and attempted;

standardized tests the student took; and honors the student received. The Transcript-level Information screen recorded the different courses as they appeared on the transcript. This information included the course title, credits and grade received for each course, the grade and the year in which each course was taken, and the different flags that indicated whether the course was an off-campus course, special education course, a course taught in English or in another language, and the level of the course (regular, honors, or remedial).

In addition to preloaded fields, the CADE system displayed labeled blank fields that the data entry clerk filled as directed. The system checked each entry to verify that it was within an allowed range and set a flag to inform the clerk when a potential error occurred. Clerks entered data exactly as it appeared on the transcript, using the Transcript Format Checklist as a guide to look for specific information on transcripts from a given school. The checklist included the student's birthdate, race/ethnicity and gender, SD/LEP status, graduation date, type of diploma awarded, details about an individual course, total number of credits received, and whether abbreviations or codes were used on the transcript. The data entry staff were instructed to use abbreviations for course titles (see exhibit 1) and to change any Roman numerals to Arabic numerals.

Exhibit 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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

The system included a type-ahead feature for the coding of course titles. As a data entry clerk entered a course title, the feature compared the letters entered against a list of course titles for the school and showed the first course title that started with the entered letters. If it was the correct course title, the data entry clerk could accept the course title. If it was the incorrect title, the clerk would continue

to type in the title. The type-ahead feature also automatically filled in known abbreviations. If a new course title was introduced to the system, it joined the list of available course titles for that particular school and became eligible for the type-ahead feature. This feature allowed for greater consistency of course titles. When all the transcripts for a school were completed, the status of the school file changed from "incomplete" to "ready for verification."

5.2.1 Verification of Transcript Data

All transcript data were 100 percent verified in the CADE system by a staff member 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 nonmatch 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 grade, year, term, course name, test name, and honors name. These six fields were displayed and reviewed by verifiers but were not key verified. For the three "name" fields, performing a visual verification rather than re-keying proved more cost-effective, as those fields were not used for any automated analyses and required the greatest number of key strokes to enter. Allowing the verifier to see the keyed course, test, or honors name also ensured that the verifier entered data in the same sequence as the original keyer.

5.3 CACE System for Coding and Editing Course Catalogs

The Computer Assisted Coding and Editing (CACE) system is a component of the MS-Access-based Data Entry and Processing system specifically created for coding high school catalogs. It consists of two major components: (1) a component for selecting and entering the most appropriate Classification of Secondary School Courses (CSSC) code and "flags" for each course in a catalog and (2) a component for matching each entry on a transcript with an entry in the corresponding school's list of course offerings. The system also provided for data selection and entry, maintained file consistency, and produced output files suitable for further analysis and manipulation. CACE's user interface was designed to reduce the likelihood of coding errors by encouraging selection from a list rather than key entry of data items.

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 was synchronized with an online version of the CSSC so that the coder could simultaneously compare the description for the course in the CSSC with the course description in the school catalog. The coder could select the appropriate CSSC code either in the suggestion list or in the corresponding section of the CSSC. If no catalog was provided, a catalog was created for the school, based on a list of courses commonly offered by high schools. The list was augmented by adding courses that reasonably would be expected to be offered, even if they did not occur on a transcript. For example, if transcripts included the first and third years of a foreign language, it was expected that the school also offered the second year of that language, even if that course did not appear on any transcript in the HSTS 2000 sample.

An alternative procedure allowed the catalog coder to 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 online CSSC or refer to the hard copy of the CSSC. 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.

5.3.1 General Procedures for Coding Course Catalogs

To ensure consistency and quality, catalog coding decisions were based on a basic set of coding principles and procedures. First, the catalog coder reviewed a school catalog "holistically" to determine the ways in which course levels, special education, and other special programs were designated. Specifically, he or she examined the sequences of courses, descriptions of programs, graduation requirements, credits awarded, and/or other available information to acquire an overview of the curriculum. Then, using CACE, the coder matched each CSSC course title with its corresponding course from the catalog, based on the available descriptions from the CSSC documentation and from the school catalog. The coder had some automated procedures to match to the CSSC. The coder could provide keywords, subject information, exact titles, or a combination of the three, and the system supplied a suggestion list of possible CSSC courses that would best match the catalog course.

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, he or she deleted those courses. Similarly, if the coder found that a course was missing from the CACE listing of catalog titles, he or she added it to the list and coded it. After the coder finished coding the regular education courses for a school, the special education expert coded all special education courses.

Figure 2 is a schematic of the data entry and coding systems illustrating the process used. The following sections describe the specific steps of the coding procedure.

5.3.2 Entering Course Titles

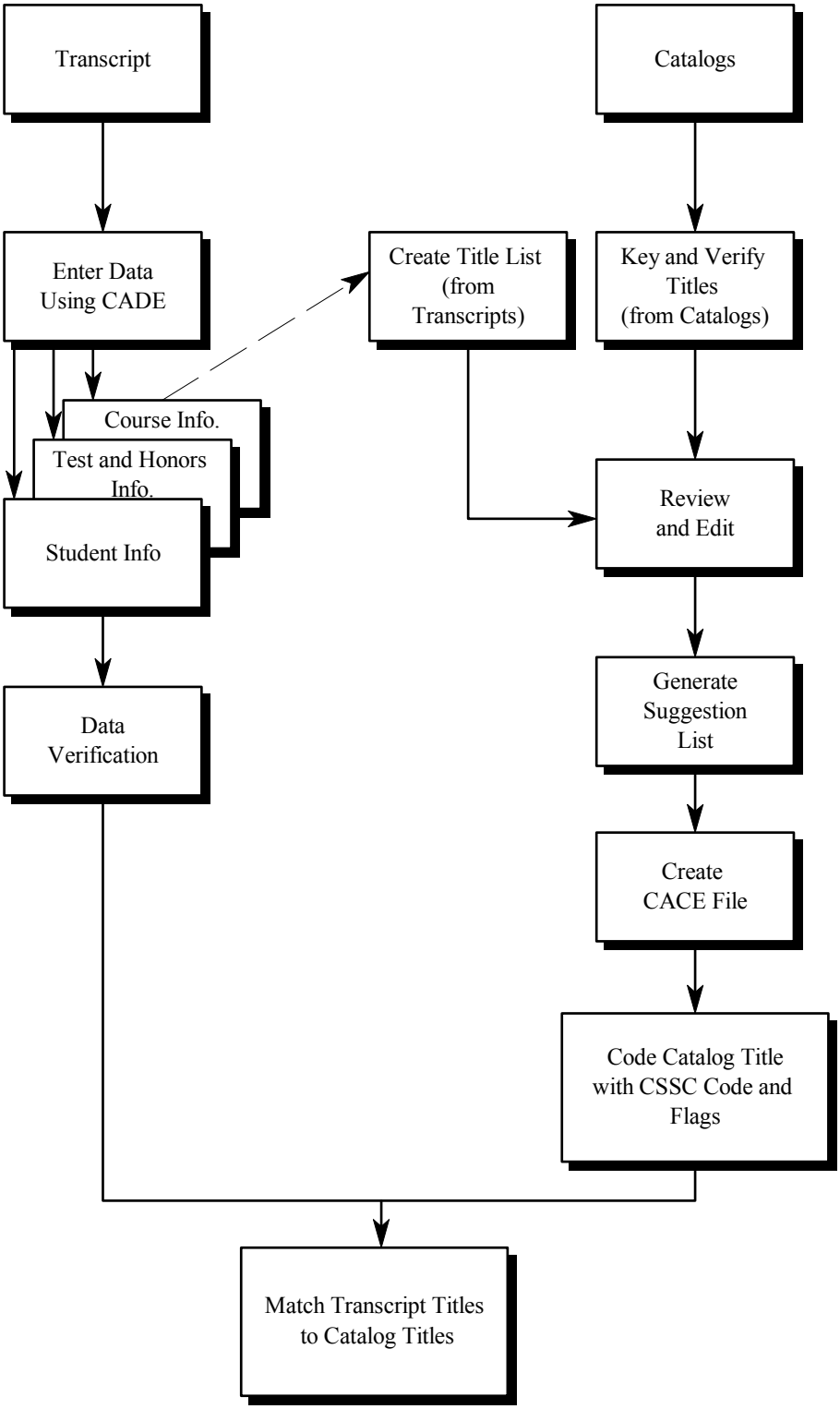
A curriculum specialist examined all catalog listings, regardless of how the catalog was created. Every attempt was made to eliminate duplication and ensure that course titles included appropriate annotations for grade ("English 10"), level ("Biology, AP"), or special programs ("Automechanics Coop Ed").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

Two variables in the School File indicate the source of information for a given school's catalog. One variable indicates whether or not the course list was derived from transcripts. The other indicates the type of catalog that the school provided (school-level catalogs or course lists, district catalogs, or schools without catalogs). To facilitate ease of use, both variables are also included in the Course Offerings File. Around 89 percent of the schools provided school level, district level catalogs or school lists.

5.3.2.1 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. An effort was made to standardize the format of titles. For example, all Roman numerals were converted to Arabic numerals. Abbreviations were standardized for all frequently appearing courses (or words in courses) such as "ADV" for "advanced," or "BEG" for "beginning," or "INTRO" for "introduction." These abbreviations were the same as those used by the transcript data entry clerks (see exhibit 1).

Figure 2. Data entry and coding process



About 69 percent of the schools provided at least one catalog, and about 11 percent provide a school list. About 75 percent of the schools provided school catalogs or school lists for two or more years. Catalogs from all years received were used to determine whether there were significant changes over the years provided. The School Information Form indicated if there were any significant changes in course offerings over the four years in which graduating students attended the school. 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 2000 attended the school. These titles were entered in the order of their appearance in the catalogs.

5.3.2.2 District-Level Catalogs

Both school-level and district-level catalogs were found at many schools. Twenty-four schools (about 9 percent) provided catalogs of courses offered by their entire school district, while the individual school's specific course offerings were a subset of those courses included in the district catalog. These district catalogs often included programs that were known not to be offered at the home school (such as an International Baccalaureate program, a vocational program, or a performing arts program). To account for courses actually offered at such schools, a list was created in the same manner as for schools not providing any catalog (i.e., creating it from titles appearing on transcripts), but the resulting list was supplemented with courses from the district catalog that were likely to be offered in the HSTS 2000 school (such as Advanced Placement English 12, Accounting, or Basic Biology) even if they did not appear on a transcript. Thus, the Course Offerings File represents the best approximation of the complete list of courses offered by the schools to their 2000 graduates in the sample.

5.3.2.3 Schools without Catalogs

Approximately 11 percent of the schools (31 of 277) did not provide any list of courses offered at the school. For these schools, which often had small student enrollments, 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. The resulting list of courses taken by students at the school was then treated as the school's catalog.

There were several limitations to creating catalogs for a school using the procedures described above. First, the list represented only the courses taken by students in the sample and might not include all courses actually offered at that school. Second, some courses had duplicates, since the same course might have been entered into the transcript file in two different formats (for example, "CONSTRUCTION 1" and "CONSTRUCTION TRADES 1" or "GLBL STDY 9" and "GLOBAL STUDIES 9"). Third, no course description was available to clarify the meaning of a title. These catalogs required considerable review and editing before course coding could proceed. To facilitate further review and edit, schools with catalogs generated using this procedure had the catalog title source variable CATSRCE set to 0 in the School File. Schools that provided catalogs or course lists had the CATSRCE variable set to 1.

5.3.3 Classification of Secondary School Courses

The Classification of Secondary School Courses (CSSC) was used as a standard for classifying and coding the courses offered by all HSTS 2000 schools and the courses appearing on all HSTS 2000 student transcripts. The CSSC is a hierarchical numbering system for all regular and special education courses offered in American high schools. Each CSSC entry includes a six-digit code, a course title and alternate titles, as well as a course description. The CSSC contains 2,268 course codes within 16 different subject areas as defined by the Secondary School Taxonomy.[8] It includes modifications made for the 1987, 1990, 1994, 1998, and 2000 HSTS. For HSTS 2000, two new codes were added to the CSSC, while five previously existing CSSC courses that were not used or duplicative were deleted. Appendix C presents more detail about the CSSC, including the entire list of CSSC codes.

The CSSC coding system employed for this purpose was a modification of the system presented in *A Classification of Secondary School Courses* (Ludwig et al. 1982). The CSSC is a modification of the college course classification system presented in *Classification of Instructional Programs* (Morgan, Hunt, and Carpenter 1991). Both course coding systems use a three-level, six-digit system for classifying courses. The CSSC uses the same first two levels as the Classification of

[8] The 16 Secondary School Taxonomy (SST) subject areas used in this study are as follows: Mathematics, Science, English, Social Studies, Fine Arts, Foreign Languages, Computer-Related Studies, Consumer and Homemaking Education, General Labor Market Preparation, Specific Labor Market Preparation, General Skills, Personal Health and Physical Education, Religion, Military Science, Special Education, and All Other Courses. The Computer-Related Studies and Special Education subject areas do not appear on the original SST. They were for HSTS research purposes.

Instructional Programs (CIP), which is represented by the first four digits of each code.[9] The third level of the CSSC (the fifth and sixth digits of the course code) is unique to the CSSC and represents specific high school courses.

A taxonomy of course subject areas was developed for the 1987 High School Transcript Study. This taxonomy, documented in the 1987 HSTS tabulations (Thorne 1988), was developed with an emphasis towards academic courses. Computer-related courses were considered as a separate non-vocational subject, and there were fewer subgroups defined for vocational and personal courses. This taxonomy was applied to data from the 1982 High School and Beyond (HS&B) First Follow-up Study and the HSTS 1987 data. The 1990 High School Transcript Study used a slightly expanded version of the same taxonomy in its reports. The 1990 study added 18 new codes to the CSSC and to the taxonomy. The full taxonomy is documented in both *The 1990 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1990, 1987, and 1982 High School Graduates* (Legum et al. 1993a) and *USER'S MANUAL: 1990 High School Transcript Study* (Legum et al. 1993c).

Starting with the 1994 study, the HSTS switched over to the Secondary School Taxonomy (SST). The SST was originally developed in 1987 under the auspices of the National Assessment of Vocational Education (NAVE) and was subject to extensive review by vocational and academic educators and researchers, NAVE staff, and contractor staff. In addition to the HS&B 1982 and HSTS 1987 files, variants of the SST were applied to files produced by the Educational Testing Service Study of Academic Prediction of Growth (1969) and the National Longitudinal Study-Youth Cohort (1975-1982), both of which were coded using unique classification schemes that were not fully compatible with the CSSC. A description of the development of the SST is provided in *The Secondary School Taxonomy Final Report* (Gifford, Hoachlander, and Tuma 1994).

Although there is broad agreement between the taxonomy developed for the HSTS 1987 and the Secondary School Taxonomy, the SST has a less purely academic emphasis and a more richly defined group of vocational education categories. Computer-related courses became vocational courses, and general skills and military science courses became new subject areas. So to maintain comparability with the earlier transcript studies, the 1987 and 1990 HSTS studies, along with the 1982 HS&B study, were recoded using the SST.

[9] Specifically, 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 students with disabilities.

The SST is limited, however, in that it contains only the CSSC codes found in the data sets which it was designed to analyze. For this reason, the SST was expanded in 1994 to include all currently defined CSSC codes.[10] The expansion of the CSSC codes led to additional changes being made to the SST for HSTS purposes. These changes did not remove any of the original SST categories, nor did they change or remove any of the CSSC codes assigned to the original SST categories. These changes were as follows:

- A second-level category called "Computer-Related Studies" was added to Academic Courses. This new category contains all CSSC codes related to computer-related studies. All the CSSC codes that appear in this new category also appear in other second-level categories, most notably the Specific Labor Market Preparation category under Vocational Courses.
- A second-level category called "Special Education" was added to Personal/Other.
- Some additional third- and fourth-level categories have been added. These new categories did not change the definition of any existing SST category. The categories were added to either further define existing categories or provide categories of educational interest.
- Drama and Dance have been separated into two categories. This split is consistent with the reporting level in the previous High School Transcript Studies. Since these two values are always reported adjacent to each other, they can easily be added together to determine the resulting combined category.

The addition of the Computer-Related Studies and Special Education categories caused some CSSC codes to be listed under two or more second-level categories. When totaling a student's overall earned credits, or credits earned in academic, vocational, and personal/other courses, these CSSC codes were only counted once. They counted toward their original SST second-level category, not the newly added category.

Because the SST assigns courses differently to academic and vocational categories than the taxonomy originally used for the HS&B 1982, HSTS 1987, and HSTS 1990, analyses based on the SST report larger numbers of students following vocational curricula and fewer numbers of students following academic curricula. Based on academic track definitions, academic program students earn at least 12 Carnegie credits in the four core academic subjects—English, social studies, mathematics, and science—

[10] In addition to the studies cited earlier in this section, the Second Follow-up of the National Education Longitudinal Study of 1988 (NELS:88) collected transcripts from high school graduates and coded them using the CSSC. The students in the transcript component of the NELS:88 study graduated from high school in 1992. Researchers at National Opinion Research Center, which conducted the NELS:88 study for NCES, were able to use the codes in the 1990 version of the CSSC and did not need to add any additional codes.

but three Carnegie credits or less in each specific labor market preparation subgroup.[11] Vocational program students earned three or more Carnegie credits in at least one specific labor market preparation subgroup, but less than 12 Carnegie credits in the four core academic courses. Using the original taxonomy developed for HSTS 1990, 69.3 percent of 1990 high school graduates were in academic programs and 7.7 percent were in vocational programs (Legum et al. 1993a). When the HSTS 1990 data were recoded using the SST, 64.1 percent of 1990 high school graduates were in academic programs and 10.4 percent were in vocational programs (Perkins et al. 2005). These changes resulted from the SST itself, and not because of any changes made for HSTS purposes.

One other feature of the SST to keep in mind is that it classifies English as a Second Language (ESL) courses as Foreign Language courses rather than as English courses. Across all HS&B and HSTS studies, this classification has the effect of lowering the number of students who satisfy the recommendation of completing 4 years of English. It also has the effect of increasing the apparent number of Foreign Language courses completed and lowering the correlations of number of years of Foreign Language completed with each set of the NAEP proficiency scores.

For the HSTS, there are two course descriptor flags associated with the CSSC: a one digit "disability" flag and a one digit "sequence" flag. The disability 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 sequence. These flags are not part of the actual CSSC code; they are included on the Master CSSC File available with the HSTS data files. The disability flag was added to the CSSC during the 1987 HSTS transcript study. The sequence flag was added during the 1990 HSTS study.

5.3.3.1 Flags

Additional information for each course was coded 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), the location at which the course was taught, and any enrollment restrictions (regular or disabled students). A full list of flags and their values is shown in exhibit 2.

[11] The eight specific labor market preparation subgroups are as follows: agriculture/renewable resources, business, marketing and distribution, health, occupational home economics, trade and industry, technical and communications, and unidentified subject.

Codes for flags were automatically set to default values when a course was selected or entered and could then be changed to nondefault values by the catalog coder. The CACE system included screens where the coder could rapidly review the flags and then edit them. The browsing screen displayed the data using one line per course title, a format that was particularly useful for locating uncoded entries and reviewing similar titles for consistency in coding flags.

Exhibit 2. Values for flags

Language Flag
0 Taught in English (DEFAULT)
1 Taught in language other than English
Off Campus Flag
0 Not an off campus course (DEFAULT)
1 Yes, taught at area Vo-Tech
2 Yes, taught at Special Ed Center
3 Yes, other
4 Yes, taught at multiple locations
Remedial/Honors Flag
1 Honors course
2 Regular course (DEFAULT)
3 Remedial course
4 International Baccalaureate
5 Advanced Placement
Sequence Flag
0 Nonsequential course (DEFAULT)
1 First course in sequence
2 Advanced course in sequence
Special Education Flag
0 Self-contained special education
1 Non special education (DEFAULT)
2 Resource-level special education
Transfer Flag
0 Not a transfer course (DEFAULT)
1 Transfer course

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

5.3.3.1.1 Coding Transfer Courses

An important variation on the course coding procedure was for transfer courses—that is, courses on a student's transcript that were taken when the student attended another school but the credits were transferred to an HSTS 2000 school and accepted there. These courses were automatically added to

the catalog list appearing in CACE with the "transfer flag" set to indicate their transfer status. In coding these transfer courses, the catalog coder used only the course title to assign CSSC codes. No descriptive information was available unless the course was taken in the same school district and a district catalog was available for review.

To address this issue, 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 in an area with a highly transient population—this automated procedure saved a great deal of time and ensured that identical titles always received identical codes.

Coders did not perform manual title matching on transfer courses. Transfer titles were automatically matched by CACE since the catalog entries were copies of transcript titles. For each transfer course, a copy of its title 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 were identical to those appearing in the transcript course list, they could be matched automatically.

5.3.3.1.2 Coding Special Education Courses

Special education courses were coded by a specialist holding an advanced degree in special education. All special education coding was reviewed by the coding supervisor, who had 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

Once the transcript data entry and verification were complete, the next step in the coding process was to match 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 both flags and course titles must be matched. For instance, "Algebra 1" with an honors flag had to be appropriately matched with an honors-level course in the catalog. For all schools, special education titles on transcripts were matched to appropriate catalog titles in special education by the supervisor.

The CACE system included 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 started 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 double-clicking the selected entry. The catalog title then appeared next to the corresponding transcript title in the upper window. This process continued until each transcript title was matched 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.

A CSSC code was assigned 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 therefore, was associated with the transcript title, based on a match of the title, course level (regular, honors, remedial), and flags (transfer, language of instruction, disability) for each transcript entry.

The matching process also served as a check on the accuracy of both transcript and catalog title data entry. For example, if an entry appeared in the transcript but not in the catalog, the catalog coder reviewed the transcript to determine whether the course should have been marked with the transfer flag. The coder also reviewed the catalog to determine whether the course was erroneously omitted from the list of catalog titles. In previous HSTS studies, this process revealed that entire programs were not described or even mentioned in the school catalog. This discrepancy occurred because the only catalog provided was out of date and different courses were offered in the graduates' high school careers than were represented in the older catalog.

One of the major difficulties encountered in evaluating transcript course titles occurred when course titles were abbreviated. The original meaning of these abbreviations was difficult to determine.

Some 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 efforts (e.g., "ARCS"). Some titles could reasonably be assigned to a broad domain, if not to a specific course. "ABC Math," for example, could be matched to the "Math-Other" course title and CSSC code. An ambiguous title was matched to an "other" course and code within a specific discipline whenever possible. Otherwise, the course was assigned a code of "60.0000" for "uncodable."

The "60.0000" CSSC code was assigned to 5,707 of the 995,035 courses entered. It represents less than 0.6 percent of the transcript entries. Note that the "60.0000" code was used to code unspecified transfer course credits; that is, when the student's transcript reported a number of transfer credits, but did not list any courses for those credits. The CSSC code was also used to add credits to those students' transcripts that had all the other attributes of a graduated senior but under 16 Carnegie credits of courses.

5.5 Standardizing Credits and Grades

Since reported credits and grade information on transcripts varied considerably among schools, districts, and states, it was necessary to standardize this information so that valid student- and school-level comparisons could be made. Standardized credit information was based on the Carnegie Unit, 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 the majority of the schools, the Carnegie Unit factor was obtained from the School Information Form as reported by the school personnel. In addition, for each school, the catalog coder filled out a Carnegie Unit Report (exhibit A-13 in appendix A). 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 Standardization of Grades table shown in exhibit A-14 in appendix A. Information was then key-entered for each school by data entry personnel. Numeric grades were converted to standardized grades as shown in table 14, unless the school documents specified other letter grade equivalents for numeric grades.

Table 14. HSTS numeric grade conversion: 2000

Numeric grade	Standard grade
90–100	A
80–89	B
70–79	C
60–69	D
< 60	F

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

5.6 Performing Quality Control Checks

As noted already, CACE had a component for selecting and entering CSSC codes and flags for courses listed in a catalog. It also matched each entry on a transcript with an entry in the school's list of course offerings. Yet another component of the CACE system automatically converted the credits on each transcript to Carnegie Units, then compared the number of credits entered to the number of credits required for graduation in that school, school district, or state (depending on which was the most reliable source of information). The number of credits required for graduation was taken from the School Information Form. This automated check verified that the total credits entered for a student were less than 150 percent of the total number of credits required for graduation and not less than 90 percent of the total credits required. This range was necessary because many students took more than the minimum requirements for graduation, while only a few students graduated with less than the required credits. When the total credits that a student had earned was either less than the number needed to graduate or more than 150 percent of the number required to graduate, the transcript and the data files were examined to see if an error had occurred. Any errors were corrected and the total credits were recalculated and compared to the graduation requirement.

The following sections describe the specific procedures used to ensure the accuracy and consistency of data entry and coding.

5.6.1 Quality Control for Transcript Data Entry

Measures to maintain the quality of data entry on transcripts included the following:

- 100 percent verification of data entry;
- 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
- reconciliation of IDs of transcripts entered with the list of valid IDs for the HSTS 2000.

Verification included all data entry fields except course titles and the term, year and grade the course was taken, test scores, and award titles.

Verification was performed by a CADE verifier who had not entered those 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 verifier a warning message if the number of credits entered was too large or small to be feasible. By reconciling the IDs that were entered on the transcripts with the IDs of students on the HSTS 2000 eligible list, it was 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 each catalog's course titles was reviewed by a curriculum specialist who visually compared the listing with the catalog. When errors were found, corrections were keyed and then the list was reviewed again. For 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

The procedures for assuring the quality of assigning CSSC codes to courses offered in HSTS 2000 schools included the following:

- careful training and supervision of coders;
- formal reporting and resolution of coding difficulties;
- reliability checking throughout the process through independent coding of a sample of courses, or by complete review of codes for nontransfer courses by the curriculum specialist;
- extensive quality reviews; and
- automated quality assurance reports.

Each of these procedures is described separately. Selection, supervision, and training of catalog coders are discussed in section 5.9. Figure 3 is a schematic diagram of the quality control procedures for catalog coding.

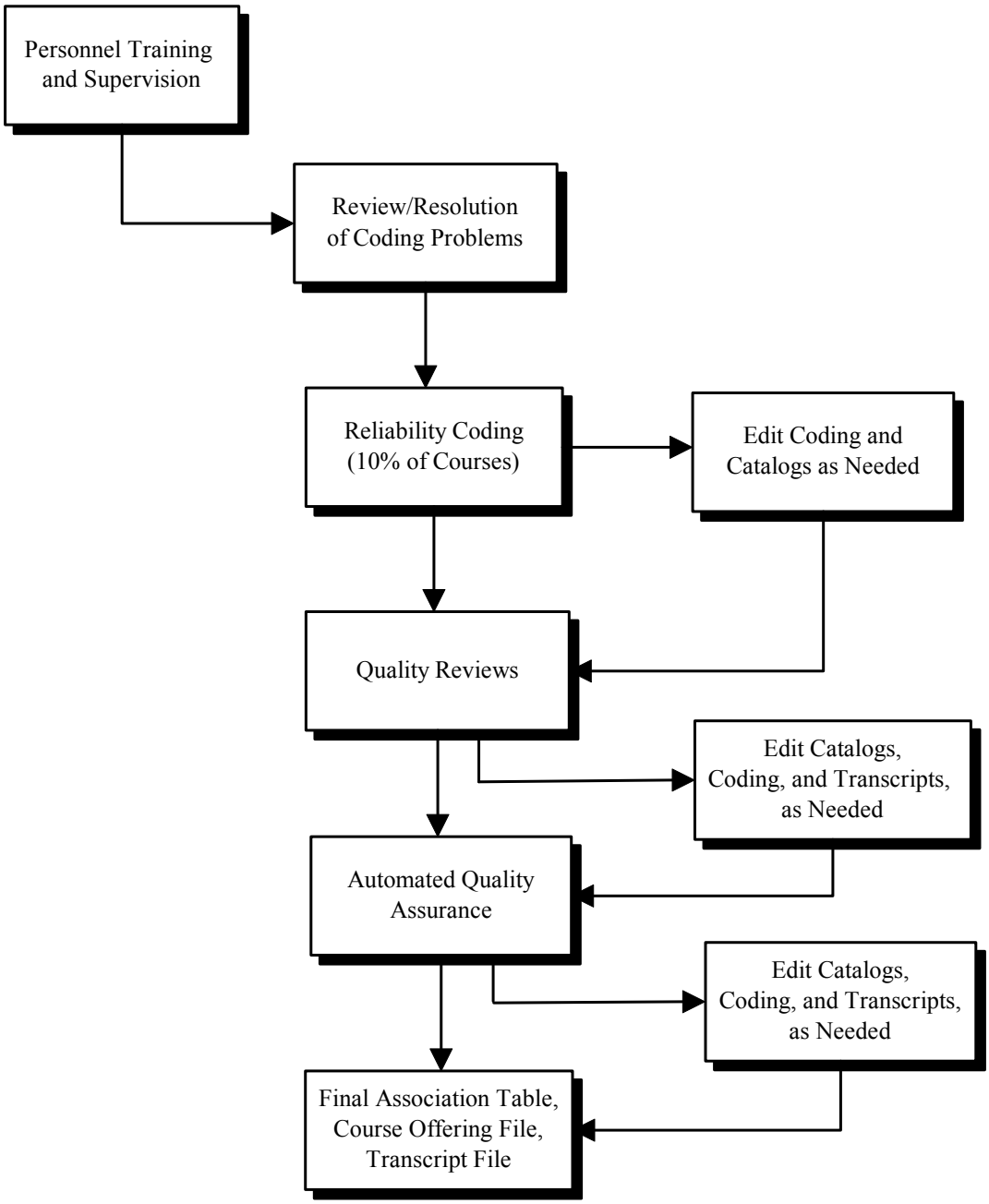
5.6.3.1 Difficulty Reporting

Problems in coding catalogs were reported directly to the curriculum specialist for review and final resolution. In some instances, additional information was obtained from school personnel to shed light on the problem encountered. Problems were resolved, and the decisions reached were documented.

5.6.3.2 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, a quality control manager coded a random sample of between 10 and 25 percent of the nontransfer courses in each school catalog.

Figure 3. Quality control processes for HSTS 2000 catalog coding



For schools with fewer than 50 nontransfer titles in their catalogs, every course was coded by the quality control manager. For schools with larger catalogs, 25 percent of the courses were coded by the quality control manager. This sample coding was then compared with the codes assigned to the same course by the catalog coder. An agreement was either an exact match of codes or a match to a code that the curriculum specialist determined was equally appropriate for the course. If 90 percent or more of the coding agreed, the quality control manager corrected the discrepancies and 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 original catalog coder, and all coding procedures and principles were reviewed, as necessary. Multiple levels of review ensured both accuracy and consistency in coding. Since all catalogs were reviewed by the coding supervisor and corrected, a high level of accuracy was achieved.

5.6.3.3 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 catalog courses that were uncoded, coded as "uncodable," or coded "other." Another report listed transcript titles that were unmatched or assigned an "uncodable" course code. The curriculum specialist reviewed all these uncoded courses and recoded and rematched to the fullest extent possible all courses for which he or 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 and catalog coding. When this review identified problems, the file was returned to a catalog coder to correct the problems, and the quality review procedures were repeated.

5.6.3.4 Automated Checks

An additional quality check took place just before the CACE files for a school were converted to an ASCII file format. Reports listing frequencies of occurrences that might indicate errors were sent to the curriculum specialist for further review. Each file was assigned one of the following status codes:

- Status 1: complete;
- Status 2: errors in transcript entry;
- Status 3: errors in catalog coding and associations; or
- Status 4: computer errors (such as duplicate course sequence numbers).

A file with a status of 2, 3, or 4 was returned to CADE and CACE for correction, a new report was generated, and the report was reviewed once more. This process was repeated until the file had a status of 1, indicating that it was complete and correct.

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.
- It was verified that all NAEP 2000 IDs in the control list also appeared on the TRF list.
- It was verified that all IDs on the TRF list for a school were in the student data file.
- A cross-tabulation of graduation year by Exit Status was created and reviewed for outliers.
- A cross-tabulation of highest year (e.g., 11th grade, 12th grade) appearing in the transcript by Exit Status was created and reviewed for outliers.
- A cross-tabulation of total Carnegie Units earned by Exit Status was created and checked for outliers.
- All students with 12th grade transfer courses (other than summer school) were listed and their transcripts checked for accuracy of data entry.
- Valid combinations of course flags were checked. For instance, no course could be both honors and remedial or special education.

5.7 Scanning and Preparing the SD/LEP Questionnaires

The SD/LEP Questionnaires (appendix B) collected during NAEP 2000 were scanned by Pearson and the files provided to the Educational Testing Service (ETS). ETS provided the HSTS 2000 with data for all 12th-grade students for whom the SD/LEP Questionnaires had been completed during NAEP 2000. Of all completed questionnaires, only the ones with corresponding records in the HSTS

2000 Student File were selected for the final HSTS 2000 SD/LEP File. A total of 2,561 students are represented in the final SD/LEP file.

The responses to the questionnaire were entered on optical scan forms by school personnel and scanned by Pearson. The data in the scanned data file were direct representations of the questionnaire responses. There were, however, seven items (questions 2, 8, 9, 10, 28, 29, and 30) on the scanned data file that needed some recoding:

- If the respondent checked a single response for the item, the value of that response was used;
- If the respondent checked two or more responses, the response code for "multiple response" was used; and
- If no response was checked, the code for "not reported" was used.

Similarly, the first item of the questionnaire, which asked for a description of the student's primary disability, was structured in such a way that allowed for multiple responses. The recoding of this item was similar as above, except that, if two or more responses were chosen, the response code for "multidisabled" was used.

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

- The TRF had the SD field flagged as 1 ("Yes");
- The student's Exit Status as entered in the CADE system was 3 or 4 (special education diploma or certificate of attendance);
- The SD/LEP Questionnaire had at least one item that was filled-out in either the SD or LEP sections.

The student's exit status, race/ethnicity, grade level, gender, birth month and year, Title I and NSLP flags were obtained from the Student File. If that information did not exist on the Student File, the corresponding data from the SD/LEP Questionnaire were incorporated if available. Frequencies and

cross-tabulations were run to check the data for valid entries and outliers before, during, and after processing.

5.8 Scanning and Preparing the School Questionnaires

The School Questionnaire was used in the NAEP 2000 and was available for 242 of the 277 HSTS 2000 schools. The data were entered on optical scan forms by school personnel and scanned by Pearson.

When processing the School Questionnaires, the system used with the previous HSTS was used. As with the SD/LEP Questionnaire, processing consisted of converting the scanned responses to provide one variable per question. When necessary, the value was set to either "multiple response" or "not reported" as appropriate. A copy of the 2000 School Questionnaire is included in appendix B.

5.9 Personnel Selection, Training, and Supervision

Trained and experienced educators were used for the coding task to ensure that coding was 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 meant. After selecting individuals with appropriate experience and background, a thorough training was conducted 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. One of the coders had served in a similar capacity for the HSTS 1998.

A curriculum specialist holding a doctorate in Curriculum and Instruction, and with experience from participation in the 1990, 1994 and 1998 High School Transcript Studies, 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 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.

5.9.1 Training Data Entry Staff

Actual transcripts were used to illustrate different formats and different types of information as demonstration materials. Trainees used these transcripts as practice exercises to gain familiarity and skill in using the CADE system. In addition, two experienced HSTS 2000 data coders prepared a summary sheet for each school that directed the data entry clerk's attention to any special features or difficulties associated with a set of transcripts.

5.9.2 Training Catalog Coders

Catalog coders who were selected had either current or prior experience teaching in American schools and/or had a college degree in education. An expert in special education was selected to code the special education courses for all schools. One of the catalog coders had coded catalogs during the HSTS 1998 and was highly experienced. He 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 four-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 operating the CACE system. Training materials included practice exercises based on actual catalogs and transcripts from HSTS 2000 schools. The first day of training consisted of classroom-type presentations 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.

6. WEIGHTING AND ESTIMATION OF SAMPLING VARIANCE

This chapter presents a detailed discussion of the weighting methodology for the 2000 High School Transcript Study (HSTS 2000). Included are sections on types of weights, adjustment procedures, and variance estimations.

The HSTS 2000 used a complex multistage sample design involving the sampling of certain subpopulations (disabled and limited English proficient (SD/LEP), Black, and Hispanic students) at higher rates. Various estimation adjustments (such as nonresponse and poststratification) were also employed to improve precision. To account for the differential sampling and various weighting adjustments, each student was assigned a sampling weight for the NAEP-linked and NAEP-non-linked populations of analysis. Sampling weights are needed to make valid inferences from the student sample to the respective populations from which they were drawn.

Sampling weights are factors assigned to each student that 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 student "represents." A student with a sampling weight of 100 represents 1 sampled student and 99 other nonsampled (or sampled but nonresponding) students in the population. A student with a sampling weight of 1 represents only the sampled student.

Two types of HSTS 2000 weights, HSTS sample weights and NAEP-linked weights, are needed for these data. The 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 HSTS NAEP-linked weights are designed for any aggregations that include only transcripts from students who were in a particular NAEP assessment. Section 6.1 discusses the weighting procedures for both types of HSTS 2000 weights.

Student estimates based on the HSTS 2000 are subject to sampling error because they are derived from a sample, rather than from the whole population. The variance is a measure of sampling error and, for the most part, determines the reliability of an estimate. Sampling variance indicates how much a population estimate for a given statistic is likely to change if it were based on another equivalent sample of individuals drawn in exactly the same manner as the achieved sample.

Since the HSTS 2000 used a complex sample design with multistage sampling, unequal selection probabilities, and complex weighting procedures, use of standard textbook formulas or standard routines in software packages such as SAS and SPSS generally underestimates the true variance for survey estimates. Instead, through the use of a variance estimation technique known as replication, replicate weights have been provided for each set of sample weights to allow users to compute variances. While there are several possible replication methods to use, HSTS 2000 replicates were derived using the stratified jackknife method, the same technique used for NAEP 2000 variances. Section 6.2 describes variance estimation procedures used for the HSTS 2000 samples.

6.1 HSTS 2000 Weighting Procedure

The High School Transcript Study provides educational policymakers and researchers with two sets of data for analyses. One set provides information regarding the course offerings and coursetaking patterns of high school graduates in the nation's secondary schools. The second set provides information on students' coursetaking patterns that can be linked to the NAEP assessment results. Each set of data requires its own set of weights to make valid inferences about the appropriate population of analyses. The HSTS sample weights are designed for all high school graduate analyses that do not involve NAEP assessment results. All students in the HSTS sample were assigned a sampling weight. The NAEP-linked weights are designed for any high school graduate analyses that involve a particular NAEP assessment. Only those students that took a NAEP assessment were assigned a linked weight.

One set of weights was generated for the HSTS 2000 sample, and four sets – one for each assessment subject (mathematics and science) and reporting population (accommodated and nonaccommodated) – were generated for the NAEP 2000 linked samples. The sets of weights were computed separately using similar weighting procedures. These procedures involved constructing a student-level weight reflecting the student's overall probability of selection and various school- and student-level weighting adjustments in order to improve precision of sample estimates. The weighting procedures for the HSTS 2000 sample weights and NAEP 2000-linked weights are described in sections 6.1.1 and 6.1.2, respectively.

6.1.1 HSTS 2000 Sample Weights

The HSTS 2000 sample weights reflect the probability-sampling scheme used to arrive at the sample of students for whom transcripts were requested. The HSTS 2000 weights were constructed without regard to the NAEP 2000 participation or nonparticipation status of schools and students. They also reflect the impact of sample nonresponse at the school and student levels, 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 (as described in Section 6.1.1.9).

6.1.1.1 Student Base Weights

The student base weight reflects a student's overall probability of being selected for the HSTS 2000. The student base weight (*STU_BWT)* may be expressed as the product

$$STU\_BWT = PSUWGT\_M \times RSCHWT \times SCH\_WT \times HSTSWT \times WIN\_WT \qquad (6.1)$$

where

- *PSUWGT_M* is the reciprocal of the probability of selection of the NAEP primary sampling unit (PSU)*;*
- *RSCHWT* is the reciprocal of the conditional probability that a given nonpublic school was included on the Private School Survey (PSS) file, given the NAEP PSU*;*
- *SCH_WT* is the reciprocal of the conditional NAEP school selection probability, given the NAEP PSU;
- *HSTSWT* is the reciprocal of the conditional HSTS school selection probability, given the NAEP PSU and NAEP school; and
- *WIN_WT* is the reciprocal of the conditional HSTS student selection probability, given the NAEP PSU and the HSTS school.

The PSU weight, *PSUWGT_M*, is the reciprocal of the probability of selection of the NAEP PSU. A total of 94 PSUs were selected for the NAEP 12th-grade sample; 22 were certainty PSUs and 72 were noncertainty PSUs. Certainty PSUs, which have 100 percent chance of selection, have a PSU weight

of 1.0. PSU weights for the noncertainty PSUs reflect probability proportional to size (PPS) sampling with one PSU per stratum.

The PSS weight, *RSCHWT*, is the reciprocal of the probability of inclusion of a nonpublic school on the PSS file, the source of the main NAEP nonpublic school frame. Public schools, which are not part of the PSS study, were assigned a PSS weight of 1.0.

The NAEP 2000 school weight, *SCH_WT*, is the reciprocal of the probability of selection of the school for NAEP conditional on the NAEP PSU.

The HSTS school weight, *HSTSWT*, is the reciprocal of the probability of selection of the school for the HSTS 2000 conditional on the NAEP PSU and the NAEP school.

The HSTS within-school student weight, *WIN_WT*, is the reciprocal of the probability of selection of the student for the HSTS 2000 conditional on the NAEP PSU and the HSTS school. If the student participated in NAEP 2000 and his/her link to NAEP was intact, this weight is the same as the NAEP within-school student weight, and it took into account the oversampling of the Black, Hispanic, and SD/LEP students where appropriate. If the student did not participate in NAEP 2000 or his/her link to NAEP was not intact, *WIN_WT* reflects the HSTS 2000 student sampling scheme described in section 3.3.

6.1.1.2 Treatment of Substitute Schools

As mentioned in chapter 2, NAEP 2000 used substitution at the school level as a way to reduce overall nonreponse. A school that replaced a refusing school (i.e., a substitute school) was assigned the school-level weighting components of the refusing school. Thus, the substitute school was treated as if it were the original school that it replaced. A substitute school was activated if its corresponding original school refused to participate. Activated substitute schools that did not participate in NAEP 2000 were effectively ignored and treated as if they never were activated. The remaining refusing original schools were adjusted for in the school nonresponse step described in section 6.1.1.4.

The 343 eligible original schools in the HSTS 2000 sample consisted of 271 schools that had participated in NAEP and 72 that did not participate. Of the 343 eligible original schools in the HSTS 2000 sample, 265 schools cooperated, resulting in an unweighted response rate of 77.3 percent. Of the 72

nonresponding original schools, 12 schools were replaced with substitutes that participated, increasing the unweighted response rate to 80.8 percent.

6.1.1.3 School Trimming Adjustment

School trimming is a weighting adjustment procedure that involves detecting and reducing extremely large school weights. Unusually large weights can seriously inflate the variance of survey estimates such as weighted means. The variability in weights contributes to the variance of an overall estimate by an approximate factor $1+V^2$, where V^2 is the relative variance of the weights. Unusually large weights are likely to produce large sampling variances of statistics of interest, especially when these large weights are associated with sample cases with rare or atypical characteristics. Weight reduction methods are typically employed to reduce the impact of these large weights on variances. The motivation behind weight reduction methods is to reduce the mean squared error of survey estimates. While the trimming of large weights reduces variances, it also introduces a small bias. However, it is presumed that the reduction in the variances outweighs the increase in the bias, thereby reducing the mean squared error.

In a number of cases, schools were assigned relatively large weights. One cause of large weights was underestimation of the number of eligible students in some schools, leading to inappropriately low probabilities of selection for those schools. A second major cause was the presence of large schools in PSUs with small selection probabilities, or large new schools in school districts with small selection probabilities. In such cases, the maximum permissible within-school sampling rate (determined by the maximum sample size allowed per school) could well be smaller than the desired overall within-PSU sampling rate.

There were several analytic approaches for detecting extremely large weights. The trimming algorithm for school weights was identical to the one used for the national main NAEP 2000 and had the effect, approximately, of trimming the weight of any school that contributed more than a specified proportion θ to the estimated variance of the estimated number of students in a given domain. Any school that contributed more than a specified proportion θ to the variance had its weight trimmed so that the school contributed exactly θ to the variance.

The following is a description of the trimming algorithm, including definitions of some variables to help in the discussion. Let

M = Number of schools in a given domain;

$SCHBWT_i$ = School base weight assigned to school "i", where

$$W_i = PSUWGT\_M_i \times RSCHWT_i \times SCH\_WT_i \times HSTSWT_i ; \tag{6.2}$$

$x_i^{'}$ = Estimated number of 12th grade students in school "i";

x_i = $SCHBWT_i \times x_i^{'}$; and

$\bar{x}$ = $(1/M)\sum_{i=1}^{M} x_i$.

The trimming criterion can be defined as follows: the x_i should satisfy

$$\forall i: \quad (x_i - \bar{x})^2 \le \theta \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ , where } \theta = 10/M . \tag{6.3}$$

This criterion can be interpreted as saying that no school should contribute more than the fixed proportion θ to the overall variance. If the initial school base weights satisfy this condition as is, then there is no trimming; i.e., the trimming adjustments SCH_TRIM_i are all set to 1.

Otherwise, the algorithm sorts the schools in descending order of $(x_i - \bar{x})^2$. This order is maintained as the x_i values are altered, so the order can be viewed as fixed even as the x_i values alter. Thus i=1 corresponds to the largest $(x_i - \bar{x})^2$, i=2 to the next largest, etc. (Note that these largest values exceeding the θ value in practice always correspond to large x_i values, so for these large values this order corresponds to a descending order by x_i.) With this re-ordering, the values of i such that

$$(x_i - \bar{x})^2 \quad \ge \quad \theta \sum_{i=1}^{M} (x_i - \bar{x})^2 \tag{6.4}$$

are i=1,…,c.

The idea behind the algorithm is to look at the x_i values as they would be if the first e records on this listing were trimmed. The e trimmed values are all equal to a fixed value x_d that satisfies the following:

$$\left(x_d - \bar{x}\right)^2 = \theta \sum_{i=1}^{M} \left(x_i - \bar{x}\right)^2 .$$

While trimming factors may vary, the x_d is invariant across all trimmed school records and the expression to the right of the equal sign may be rewritten as

$$= \theta e \left(x_d - \bar{x}\right)^2 + \theta \sum_{i \notin A} \left(x_i - \bar{x}\right)^2 , \tag{6.5}$$

where e is the number of records trimmed and A is the set of records trimmed. Gathering all terms to the left of the equal sign results in:

$$(1 - \theta e)\left(x_d - \bar{x}\right)^2 - \theta \sum_{i \notin A} \left(x_i - \bar{x}\right)^2 = 0 , \tag{6.6}$$

which may be rewritten as

$$\left(1 - \theta e\right) \left[\frac{\left(M - e\right) x_d - \sum_{i \notin A} x_i}{M} \right]^2 - \theta \sum_{i \notin A} \left[\frac{M x_i - e x_d - \sum_{j \notin A} x_j}{M} \right]^2 = 0 . \tag{6.7}$$

After expanding the squared expressions, this becomes

$$\begin{aligned} &\frac{\left(1 - \theta e\right)}{M^2} \left[\left(M - e\right)^2 x_d^2 - 2\left(M - e\right) \left(\sum_{i \notin A} x_i \right) x_d + \left(\sum_{i \notin A} x_i \right)^2 \right] \\ &- \frac{\theta}{M^2} \sum_{i \notin A} \left[e^2 x_d^2 - 2e \left(M x_i - \sum_{j \notin A} x_j \right) x_d + \left(M x_i - \sum_{j \notin A} x_j \right)^2 \right] = 0 \end{aligned} , \tag{6.8}$$

which can be rewritten as a quadratic equation in x_d

$$\left[(1-\theta e)(M-e)^2-\theta(M-e)e^2\right]x_d^2$$
$$+\left[(1-\theta e)(-2)(M-e)\sum_{i\notin A}x_i+(\theta)(2e^2)\sum_{i\notin A}x_i\right]x_d\,. \qquad (6.9)$$
$$+\left[(1-\theta e)\left(\sum_{i\notin A}x_i\right)^2-\theta\sum_{i\notin A}\left(Mx_i-\sum_{j\notin A}x_j\right)^2\right]=0$$

This further simplifies to:

$$(M-e)(M-e-\theta eM)x_d^2 \;-\; 2(M-e-\theta eM)\left(\sum_{i\notin A}x_i\right)x_d \;-\; \theta M^2\sum_{i\notin A}x_i^2$$
$$+\quad (1+\theta M)\left(\sum_{i\notin A}x_i\right)^2 \;=\; 0 \qquad (6.10)$$

Substituting $10/M$ for θ in the above expression gives

$$(M-e)(M-11e)x_d^2-2(M-11e)\left(\sum_{i\notin A}x_i\right)x_d-10M\sum_{i\notin A}x_i^2+11\left(\sum_{i\notin A}x_i\right)^2=0\,. \qquad (6.11)$$

Solving for x_d produces this simple expression:

$$x_d=\bar{x}+\sqrt{\left(\frac{10M}{M-11e}\right)\sigma_x^2} \qquad (6.12)$$

where $\bar{x}=\dfrac{\sum_{i\notin A}x_i}{M-e}$ is the mean of the x_i among untrimmed school records, and $\sigma_x^2=\dfrac{\sum_{i\notin A}x_i^2}{M-e}-\left(\dfrac{\sum_{i\notin A}x_i}{M-e}\right)^2$ is the variance of the x_i among untrimmed school records.

The critical problem that led to the use of an iterative process in the past is that, after trimming *e* records and assigning a new x_d to these records, the recomputed $\bar{x}$ and sum of squares may indicate that further records (e.g., record *e*+1) may now violate the trimming criterion. Under this procedure, an x_d is generated according to these formulas for each and every potential value of *e*, going

down the sorted list in a single step. The correct final value of e is the first e value for which the recomputed x_i's, the sum of squares, and proportions of sum of squares all satisfy the trimming criterion.

The trimming procedure for public schools was done within each NAEP region. For nonpublic schools, the trimming procedure was done by Catholic/non-Catholic status. The outcome of the trimming procedure was that one public school was trimmed in the West NAEP region with a trimming factor of 0.52325.

6.1.1.4 School Nonresponse Adjustment

The school nonresponse adjustment procedure inflates the weights of schools that participated in the HSTS 2000 to account for eligible schools that did not participate. School nonresponse leads to the loss of sample data that must be compensated for in the weights. Similar to the school trimming procedure, the purpose of the nonresponse adjustment procedure is to reduce the mean square error of survey estimates. While the nonresponse adjustment reduces the bias from the loss of sample, it also increases variability among the survey weights leading to increased variances. However, it is presumed that the reduction in bias outweighs the increase in the variance, thereby reducing the mean squared error of survey estimates.

As mentioned in section 6.1.1.2, substitute schools were used as a step to reduce nonresponse at the school level. A cooperating substitute school took the place of its corresponding refusing original school and was used in the subsequent student sampling stage. The school nonresponse adjustment was used to adjust for the remaining school that did not cooperate even after the use of substitution.

6.1.1.4.1 Determining School Nonresponse Cells

School-level nonresponse cells for the HSTS 2000 were determined based on the quasi-randomization approach (Oh and Scheuren 1983), where nonresponse cells are defined using school characteristics known to be related to response. Every school in the sample was assigned to a nonresponse cell based on its characteristics. The critical assumption under the quasi-randomization approach was that the response rate was homogeneous within the nonresponse cells. This approach was implemented using a

classification algorithm known as CHAID (Chi-square Automatic Interaction Detector). CHAID divides a population into homogeneous subgroups with respect to a target characteristic. In the case for defining nonresponse cells, the target characteristic was response propensity.

CHAID analysis was used to determine the nonresponse cells for the HSTS 2000 (Lee et al. 1989). The CHAID analysis began by dividing the population into two or more groups based on categories of the best response propensity predictor. Each group was then divided into smaller subgroups based on the best available predictor at each level. The splitting process continued until either there was no significant predictor remaining or the minimum cell size requirement was met. The CHAID software displayed the final subgroups in the form of a tree diagram whose branches (nodes) corresponded to the subgroups.

The HSTS 2000 used four school- or PSU-level variables in determining nonresponse cells. Nonresponse cells were defined based on cross-classifications of these PSU characteristics, and each was required to have a minimum of six cooperating schools. The four variables used to define nonresponse cells were:

1. Metropolitan/nonmetropolitan PSU status;
2. NAEP region;
3. Public/nonpublic school status; and
4. High minority status: whether or not the school had greater than 15 percent minority students

CHAID produced eight nonresponse cells as shown in table 16 in section 6.1.1.4.2.

6.1.1.4.2 School Nonresponse Adjustment Factors

In each nonresponse cell *h*, the nonresponse adjustment factor was calculated as follows:

$$SCH\_NRAF_h = \frac{\sum_{B_h} SCHBWT_{h,i} \times SCH\_TRIM_{h,i}}{\sum_{C_h} SCHBWT_{h,i} \times SCH\_TRIM_{h,i}} \quad (6.13)$$

where

- $SCHBWT_{h,I}$ = the school base weight for school i in nonresponse cell h;
- $SCH\_TRIM_{h,i}$ = the school trimming factor for school i in nonresponse cell h;
- B_h = the set of all eligible HSTS schools in nonresponse cell h; and
- C_h = the set of all eligible schools in nonresponse cell h that participated in the HSTS.

As mentioned earlier, a participating substitute school took the place of a nonparticipating original school and took on the weighting components of the original school that it replaced.

Table 15 shows the definitions of the eight nonresponse cells, the number of HSTS 2000 participating schools, and the school nonresponse adjustment factor for each cell.

Table 15. School nonresponse adjustment factors for the HSTS sample weights: 2000

School nonresponse cell	Number of HSTS 2000 participating schools	School nonresponse adjustment factors (SCH_NRAF)
Public—NonMSA schools		
Northeast	7	1.2422
Southeast and Central	41	1.0649
West	18	1.0000
Public—MSA schools		
Northeast	35	1.4011
Southeast	44	1.0909
Central	32	1.4335
West	72	1.2144
Nonpublic		
All regions	28	1.3040

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

Each participating school in a nonresponse cell received the same nonresponse adjustment factor corresponding to that cell.

6.1.1.5 School Substitution Adjustment

The substitution adjustment factor adjusted for the difference in grade enrollment prior to sampling between the participating substitute school and its corresponding original school that it replaced. It applied only to participating substitute schools and was calculated as follows:

$$SUBADJ = \frac{GRD\_ENRL}{SUB\_ENRL} \tag{6.14}$$

where *GRD_ENRL* was the grade enrollment of the original school and *SUB_ENRL* was the grade enrollment of its corresponding substitute school.

This adjustment was necessary because the weights of students in a substitute school should reflect what the weights of the students in its corresponding original school would have been if the original school participated. *GRD_ENRL* and *SUB_ENRL* represented the best estimate of the number of students eligible for sampling in the original and substitute school, respectively.

To illustrate the purpose of this adjustment, suppose an original school that did not participate had 300 students enrolled in 12th grade and the substitute school that replaced it had 200 students in the 12th grade. The sample of 100 students from the substitute school each had a within-school weight of 2.0 (= 200/100). However, if the original school had participated, each of the 100 students in the sample would have had a within-school weight of 3.0 (= 300/100). The substitution adjustment factor of 300/200 was applied to the students in the substitute school so that the adjusted within-school weight was 3.0—that is, 2.0 × (300/200) = 3.0.

There were 12 substitute schools in the HSTS 2000 sample that required the substitution adjustment factors. All students in these schools received the factor corresponding to their school. Table 16 shows the distribution of the substitution adjustment factors across the 12 substitute schools. All students in the original schools received a *SUBADJ* factor of 1.0.

Table 16. Distribution of substitution adjustment factors for the HSTS sample: 2000

Distribution	Substitution adjustment factors
Minimum	0.6511
25th percentile	0.8559
50th (median)	1.0242
75th percentile	1.1600
Maximum	1.4234
Mean	1.0175

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.1.6 Year-Round School Weighting Adjustment

The year-round adjustment factor applied only to students in year-round schools, where only a portion of the total student body was in school at any given point in time. The year-round adjustment factor inflated the weight to account for students who were on break at the time of student sampling. It was calculated as

$$YRRND\_FC = \frac{1}{1 - PCT\_OFF}, \tag{6.15}$$

where *PCT_OFF* was the percentage of students that were on break at the time of student sampling. Only 2 of the 277 cooperating schools were year-round schools. In both, 33 percent of the students were off track at the time of sampling. Each student in these two schools received a year-round factor of 1.4925. The students not in year-round schools received a factor of 1.0.

6.1.1.7 HSTS Student Nonresponse Adjustment

The HSTS student nonresponse adjustment procedure inflated the weights of "responding" students to account for "nonresponding" eligible students. Students who graduated in the 2000 school year were considered respondents (students with transcripts) if transcripts were received for at least 75 percent of the credits required by their school to graduate, and nonrespondents (i.e., students with missing

transcripts) otherwise. An exception was made for graduating students with a special education degree. They were considered respondents as long as at least one transcript was received. Students who did not graduate in school year 2000 were considered ineligible for the HSTS 2000, regardless of whether transcripts were received. They were not included in this adjustment but were retained since they were needed for the poststratification process.

Student nonresponse adjustment factors compensate the weights for the loss of data associated with the eligible students with missing transcripts. As with its counterpart at the school level, the student nonresponse adjustment was intended to reduce the mean squared error of the HSTS 2000 estimates.

Student nonresponse adjustment cells were determined based on the same approach as the school nonresponse adjustment using a CHAID analysis. The potential variables used to define the student nonresponse cells were the following:

1. Type of degree (standard degree, honors degree, special education certificate, certificate of attendance, or certificate of completion);
2. Age classification (born before 10/79, born on or after 10/79);
3. Race classification (1:White or Asian Pacific Islander; 2:Black, Hispanic, American Indian, or other race);
4. NAEP region (Northeast, Southeast, Central, West);
5. Gender (male, female);
6. School type (public, nonpublic);
7. Metro status (metropolitan, nonmetropolitan).

The minimum cell requirement was 30 responding students. Table 17 shows the definitions of the final student nonresponse cells.

Table 17. Student nonresponse adjustment cells and factors for HSTS weights: 2000

Cell number	Student nonresponse cell	Number of HSTS 2000 responding students	Student nonresponse adjustment factors
1	Standard degree; older age group; White or Asian Pacific Islander	2,518	1.0079
2	Standard degree; older age group; Black, Hispanic, American Indian, or other race	1,668	1.0119
3	Standard degree; younger age group; Northeast region; nonmetro area; male	136	1.0145
4	Standard degree; younger age group; Northeast region; nonmetro area; female	159	1.0000
5	Standard degree; younger age group; Northeast region; metro area	2,686	1.0028
6	Standard degree; younger age group; Southeast or Central region	6,265	1.0044
7	Standard degree; younger age group; West region; nonmetro area	784	1.0000
8	Standard degree; younger age group; West region; metro area; public school	4,570	1.0097
9	Standard degree; younger age group; West region; metro area; nonpublic school	73	1.0000

See notes at end of table.

Table 17. Student nonresponse adjustment cells and factors for HSTS weights: 2000—Continued

Cell number	Student nonresponse cell	Number of HSTS 2000 responding students	Student nonresponse adjustment factors
10	Honors degree; White or Asian Pacific Islander; nonmetro area	158	1.0065
11	Honors degree; White or Asian Pacific Islander; metro area	959	1.0000
12	Honors degree; Black, Hispanic, American Indian, or other race; older age group	56	1.0106
13	Honors degree; Black, Hispanic, American Indian, or other race; younger age group	336	1.0024
14	Special Ed or certificate of completion; Northeast or Southeast region; older age group; White or Asian Pacific Islander	69	1.0154
15	Special Ed or certificate of completion; Northeast or Southeast region; older age group; Black, Hispanic, American Indian, or other race	80	1.0467
16	Special Ed or certificate of completion; Northeast or Southeast region; younger age group	99	1.0854
17	Special Ed or certificate of completion; Central or West region; White or Asian Pacific Islander	59	1.0281
18	Special Ed or certificate of completion; Central or West region; Black, Hispanic, American Indian, or other race	68	1.0000
19	Certificate of attendance; White or Asian Pacific Islander	53	1.0178
20	Certificate of attendance; Black, Hispanic, American Indian, or other race	135	1.0091

NOTE: "Older age group" is defined as born before 10/81 and "younger age group" otherwise.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

In each nonresponse adjustment cell h, the student nonresponse adjustment factor STU_NRAF_h was calculated as follows:

$$STU\_NRAF_h = \frac{\sum_{B_h} STU\_BWT_{h,i} \times SCH\_TRIM_{h,i} \times SCH\_NRAF_{h,i} \times SUBADJ_{h,i} \times YRRND\_FC_{h,i}}{\sum_{C_h} STU\_BWT_{h,i} \times SCH\_TRIM_{h,i} \times SCH\_NRAF_{h,i} \times SUBADJ_{h,i} \times YRRND\_FC_{h,i}} \quad (6.16)$$

where

- B_h = the set of eligible HSTS 2000 students (i.e., graduates with or without transcripts) in student nonresponse cell h;
- C_h = the set of graduates with complete and usable transcripts (i.e., respondents) in student nonresponse cell h;
- $STU\_BWT_{h,i}$ = the student base weight, as defined in section 6.1.1.1, for student i in student nonresponse cell h;
- $SCH\_TRIM_{h,i}$ = the school trimming adjustment, as defined in section 6.1.1.3, for student i in student nonresponse cell h;
- $SCH\_NRAF_{h,i}$ = the school nonresponse adjustment factor, as defined in section 6.1.1.4, for student i in student nonresponse cell h;
- $SUBADJ_{h,i}$ = the school substitution adjustment, as defined in section 6.1.1.5, for student i in student nonresponse cell h; and
- $YRRND\_FC_{h,i}$ = the year-round school weighting adjustment, as defined in section 6.1.1.6, for student i in student nonresponse cell h.

The HSTS 2000 sample weight used in the student nonresponse procedure reflected the student base weight and all school- and student-level weighting adjustments prior to this adjustment.

Each graduate with complete and usable transcripts (i.e., respondent) in a nonresponse cell received a nonresponse adjustment factor, as calculated above, corresponding to that cell. Ineligible HSTS 2000 students (i.e., those who did not graduate) received a nonresponse adjustment factor of 1.0. Table 17 summarizes the student nonresponse cells along with the number of responding students and the nonresponse adjustment factors for each cell.

6.1.1.8 Student Trimming Adjustment

Another trimming adjustment procedure was done to detect and trim extremely large weights at the student level. Large student weights generally resulted from compounding nonresponse adjustments at the school and student levels coupled with low to moderate probabilities of selection at the various stages of sampling. As with school trimming weights, the purpose of the trimming student weights was to reduce the effect of unusually large weights on survey estimates. Trimming may introduce a small bias but is designed to reduce the mean square error of sample estimates.

The trimming algorithm was identical to that used for the main NAEP 2000 survey. The algorithm had the effect of trimming the overall weight of any school that contributed more than a specified proportion θ to the estimated variance of the estimated number of students eligible for the HSTS 2000 survey.

The student trimming adjustment worked similarly to the school trimming adjustment. Student weights (through the student nonresponse adjustment) were trimmed if their school contributed more than a specified proportion to the variance on the estimated number of students within a given domain. School-level estimates of students were calculated by summing the weights of students (i.e., graduates and nongraduates) in the school. See the description of the school trimming procedure in section 6.1.1.3.

The domains used for student trimming were NAEP region for public schools and Catholic/non-Catholic for nonpublic schools, the same as in the school trimming procedure. None of the student weights needed trimming. This can be attributed to the high response rate at the student level (about 97% overall). High response rates mean that little to no adjustment was needed for student nonresponse.

6.1.1.9 Poststratification Adjustment

Poststratification is a weighting procedure that adjusts the weights of sample cases so that the weighted sample distribution is the same as some known population distribution. That is, the sums of the poststratified-adjusted weights are equal to known population totals for certain subgroups of the population. The main purposes of poststratification are to improve precision of survey estimates by

reducing their mean squared error and to enhance the comparability of survey data with other surveys, particularly when comparing estimates from the same survey over time.

The poststratification adjustment procedure used for the HSTS 2000 involved applying a ratio adjustment to student weights. Eligible and ineligible HSTS 2000 students were partitioned into poststratification cells, and a single ratio adjustment factor was calculated and applied to the weights of all students in a given cell. The numerator of the poststratification factor was an independent estimate of the number of students in the given cell, and the denominator was the corresponding estimate derived using the HSTS sample weights. The numerator was derived from 1997 and 1998 Current Population Survey (CPS) data and 1999 population projections made by the U.S. Census Bureau. (Details of the method used to derive these independent estimates are given in the forthcoming online 2000 NAEP Technical Report.[11]) Ineligible students were included in the poststratification adjustment because CPS totals do not distinguish between graduates and nongraduates.

Poststratification adjustment cells were defined in terms of race, ethnicity, and region for 12th-grade students 17 years old or younger. Students 18 years or older were not included in the poststratification because it is not possible to derive reliable counts from the CPS data. The CPS counts all adult education students, regardless of age, as 12th-grade students.

The poststratification factor for student i in a given poststratification adjustment class h was calculated as follows:

$$RPTPS\_AD_h = \frac{TOTAL_h}{\sum_{C_h} STU\_TRMWT_{h,i}} \tag{6.17}$$

where

- $TOTAL_h$ = the total number of 12th grade students 17 years old or younger in poststratification cell h from Census Bureau data;
- $STU\_TRMWT_{h,i}$ = the HSTS sample weight through the student trimming procedure for student i in poststratification cell h; and
- Set C_h = the set of eligible and ineligible 12th grade students 17 years old or younger in the HSTS 2000 sample in poststratification cell h.

[11] At the time of publication, the online NAEP 2000 Technical Report did not yet have an official web address. When published, information on its location can be found at the National Assessment of Educational Progress web site (http://nces.ed.gov/nationsreportcard).

The HSTS sample weight through the student trimming procedure reflected the student base weight and all school- and student-level weighting adjustments prior to poststratification. Note that students who were age 18 or older received the poststratification factor according to their poststratification cell, even though they were not used in calculating the factor.

Table 18 provides the control totals, sum of the unpoststratified weights, and poststratification adjustment factors for each poststratification cell.

Table 18. Poststratification adjustments for the HSTS sample: 2000

Post-stratification cell	Race/ethnicity	Region	Age group	Post-stratification adjustment	Control total	Sum of unpoststratified weights
1	Black	All	≤ 17	1.2359	334,181	270,400
2	Hispanic	All	≤ 17	1.1908	275,294	231,190
3	Other race	All	≤ 17	0.9473	136,643	144,248
4	Non-Hispanic White	Northeast	≤ 17	1.0836	362,426	334,463
5	Non-Hispanic White	Midwest	≤ 17	1.2736	519,392	407,825
6	Non-Hispanic White	South	≤ 17	1.1683	598,515	512,315
7	Non-Hispanic White	West	≤ 17	1.3444	359,390	267,318

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.1.10 Final HSTS Sample Weight

The HSTS 2000 sample estimates of transcript characteristics were based on poststratified student-level weights assigned to high school graduates with usable transcripts. High school graduates with missing transcripts (i.e., HSTS 2000 nonrespondents) and nongraduates (i.e., HSTS 2000 ineligibles) were not included in the data set since they were not used in forming sample estimates. The final student weight reflects the student base weight and various school- and student-level weighting adjustments. The final student-level HSTS sample weight, *FSTUWT*, is given by

$$FSTUWT = STU\_BWT \times SCH\_TRIM \times SCH\_NRAF \times SUBADJ \times YRRND\_FC \times STU\_NRAF \times STU\_TRIM \times RPTPS\_AD \quad (6.18)$$

where

- *STU_BWT* = Student base weight (as defined in section 6.1.1.1);
- *SCH_TRIM* = School trimming adjustment factor (as defined in section 6.1.1.3);
- *SCH_NRAF* = School nonresponse adjustment factor (as defined in section 6.1.1.4);
- *SUBADJ* = School substitution adjustment factor (as defined in section 6.1.1.5);
- *YRRND_FC* = Year-round school adjustment factor (as defined in section 6.1.1.6);
- *STU_NRAF* = Student nonresponse adjustment factor (as defined in section 6.1.1.7);
- *STU_TRIM* = Student trimming adjustment factor (as defined in section 6.1.1.8); and
- *RPTPS_AD* = Poststratification factor (as defined in section 6.1.1.9).

The distribution of the final student weight for the HSTS 2000 sample is given in table 20.

Table 19. Distribution of final HSTS student weights: 2000

Sample distribution	HSTS sample weights
Number of graduates with transcripts	20,931
Student weights	
Total	3,012,000
Minimum	23.4990
25th percentile	83.3203
50th percentile (median)	111.9566
75th percentile	200.5194
Maximum	904.3071
Mean	143.9148

NOTE: The coefficient of variation measures the spread of a set of data as a proportion of its mean. This percentage is the ratio of the sample standard deviation to the sample mean.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.2 HSTS NAEP-Linked Weights

The HSTS NAEP-linked weights allow users to analyze the relationship between students' proficiencies, as measured by their outcomes on the NAEP 2000 assessments, and students' coursetaking in their high school careers. Twelfth-grade students in these populations of analyses participated in a

given NAEP 2000 assessment, have a completed transcript, and graduated as determined by the HSTS 2000. There were 1,413 students for whom a completed transcript was received but no NAEP 2000 assessment existed (because either the school or the student refused to participate in NAEP 2000 or the student was absent on assessment day). These students can be part of the HSTS 2000 database but not the linked database that requires both transcripts and assessment results for the same student.

The students in the linked database required a different set of sampling weights than those in the HSTS 2000 database alone, as the set of students that qualified for these databases was a subset of the larger HSTS 2000 set. In particular, the school and student nonresponse adjustments are larger for the linked weights than for the HSTS 2000 weights. This difference is so because a student or school had to participate in both the NAEP 2000 and the HSTS 2000 surveys to qualify as a "respondent" for the linked database. The schools also had to maintain the link between the HSTS transcripts and NAEP student assessment scores. This requirement reduced the number of school and student responses, thereby increasing the nonresponse adjustment factors. Table 21 shows the number of schools and students in the HSTS 2000 sample by HSTS/NAEP response status.

Table 20. Total number of schools and students in the HSTS by HSTS/NAEP response status: 2000

Response status	Number of sampled schools	Number of sampled students
Total	277	23,522
HSTS and NAEP cooperating schools, with linkage	248	22,010
HSTS cooperating, but not NAEP	16	662
HSTS cooperating, no NAEP link	13	850

NOTE: The number of schools includes original and substitute schools.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

Four sets of NAEP-linked weights were computed, one for each assessment (i.e., mathematics and science) by reporting population (i.e., accommodated and nonaccommodated) sample. The linked weights were computed using a weighting procedure similar to the HSTS sample weights. Each assessment/reporting population sample represented the full population, so each of the four sets of NAEP-linked weights aggregated separately to the population totals. This section of the report describes the weighting procedure used to weight the NAEP-linked samples.

Defining reporting populations for the NAEP-linked weights require understanding the NAEP 2000 school sample types. As explained in section 2.4, two different sample types were assigned to schools. In sample type 3 (S3) schools, accommodations were offered to students with disabilities (SD) and students with limited English proficiency (LEP). In sample type 2 (S2) schools, no assessment accommodations were offered to SD/LEP students.

6.1.2.1 Reporting Populations

The HSTS NAEP-linked data can be analyzed for two types of reporting populations. One is the NAEP 2000 reporting population made up of nonaccommodated students. This population was represented by students who did not need accommodations for a given NAEP assessment (non-SD and non-LEP students in either S2 or S3 schools) along with SD or LEP students who were assessed under the set of administration rules that did not provide accommodations (i.e., in S2 schools). The nonaccommodated reporting population is referred to as the "R2" reporting population in this section of the report.

The second reporting population consisted of the accommodated students, which become the standard reporting population beginning with the NAEP 2002 assessments. It was represented by students who did not need accommodations (i.e., non-SD and non-LEP students in either S2 or S3 schools), along with SD or LEP students who were assessed under the set of administration rules that provided accommodations (i.e., in S3 schools). The reporting population of accommodated students is referred to as the "R3" reporting population in this section of the report.

For practical reasons, final student weights for each reporting population were derived together according to the steps described below except for the poststratification adjustment. As mentioned above, every non-SD and non-LEP student was in both reporting populations. With roughly 90 percent of the total 12th-grade sample in both samples, weighting adjustments done separately by reporting population would have yielded little, if any, difference in the adjustment factors. Poststratification was done separately by reporting population since each set of weights is to sum up to population totals.

6.1.2.2 Student Base Weights

The student base weight reflected a student's overall probability of being selected for the given HSTS NAEP-linked sample. It was the product of the HSTS 2000 student base weight and four factors related to NAEP sampling:

1. the conditional probability, given the sample of schools in a PSU, that the school was assigned a specific assessment subject;
2. the probability that the school was assigned a particular sample type that dictated the set of administration rules used for the assessment;
3. student-level subject allocation weighting factor; and
4. reporting population subsampling factor.

Thus, the NAEP 2000-linked student base weight ($STU\_BWT_{r,s}$) for a given assessment subject s and reporting sample r may be expressed as the product

$$STU\_BWT_{r,s} = STU\_BWT_{HSTS} \times ASBJWT_s \times STYWT \times SPL_s \times REP\_FCTR_r \qquad (6.19)$$

STU_BWT_{HSTS} was the student base weight for the HSTS 2000 sample as described in section 6.1.1.1. It reflected a student's overall probability of selection for the HSTS 2000.

The subject assignment weight, $ASBJWT_s$, is the reciprocal of the probability that the particular subject s was assigned to the school. Subject assignment weights varied by number of subjects assigned to a school, subject of the assessment, and school type (public or nonpublic). If either a public or nonpublic school was large enough to assess both subjects, then both subjects were assigned to that school. Students in that school in either mathematics or science received a weighting factor of 1.0. If a school was large enough for only one subject, then one half of nonpublic schools were assigned mathematics and the other half science, while 7/16 of public schools were assigned mathematics and 9/16 were assigned science. Students assigned to either mathematics or science in such nonpublic schools received a weighting factor of 2.0. Students in such public schools assigned mathematics received a factor of 16/7 and those assigned to science received a factor of 16/9. Table 21 summarizes the subject assignment weight by subject, school type, and the number of subjects assigned to a school.

Table 21. Subject assignment weights ($ASBJWT_s$) by school type and assessment: 2000

	Public schools		Nonpublic schools	
Number of subjects	Mathematics	Science	Mathematics	Science
1	16/7	16/9	2.0	2.0
2	1.0	1.0	1.0	1.0

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

The sample type weight, *STYWT*, reflects the splitting of the 12th-grade school sample into two equal-size subsets to enable Educational Testing Service (ETS) to analyze two different sets of administration rules (one with accommodations and the other without accommodations). Each school was assigned a sample type weight of 2.0.

The subject allocation weighting factor, SPL_s, adjusts for allocating sampled students across the subjects assigned to a school. It differs from the subject assignment weight $ASBJWT_s$ because it reflects the assignment of subjects to students instead of schools. If a school was assigned only one subject, then all students were assigned to that subject and received a weighting factor of 1.0. If a school was assigned two subjects, the sample allocation varied by public and nonpublic. For public schools, 7/16 of the students were assigned to mathematics and 9/16 of the students were assigned to science. Thus, students assigned to mathematics and science received respective weighting factors of 16/7 and 16/9. For nonpublic schools, half the students were assigned to mathematics and the other half to science, and so each student received a weighting factor of 2.0. Table 22 summarizes the subject allocation weights.

Table 22. Student-level subject allocation weights (SPL_s) by school type and assessment: 2000

	Public schools		Nonpublic schools	
Number of subjects	Mathematics	Science	Mathematics	Science
1	1.0	1.0	1.0	1.0
2	16/7	16/9	2.0	2.0

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

The reporting factors, RPT_FCTR_r, assigned to students were specific to the reporting populations but did not vary by subject. Each assessed and excluded student in a specific reporting population received a reporting factor as shown in table 23.

Table 23. Reporting factors (RPT_FCTR_r) by reporting population, sample type, and SD and LEP status: 2000

	R2 Reporting population		R3 Reporting population	
Sample type	Non-SD/LEP	SD/LEP	Non-SD/LEP	SD/LEP
S2	0.5	1.0	0.5	†
S3	0.5	†	0.5	1.0

† Not applicable
NOTE: The S2 sample type indicates schools that did not offer accommodations to SD/LEP students, while the S3 sample type indicates schools that did offer accommodations to SD/LEP students.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.2.3 Treatment of Substitute Schools

Similar to the HSTS 2000 sample weighting, a participating substitute school that took the place of a refusing school was assigned the weighting components of that refusing school. The subject assignment weight ($ASBJWT_s$) for a substitute school was adjusted appropriately if it was not large enough to assess all sessions or subjects assigned to the original school. Among the 12 substitute schools in the HSTS 2000 sample schools, 11 substitute schools participated in both the mathematics and science assessments, while the other substitute schools participated only in the science assessment.

6.1.2.4 School Trimming Adjustment

The same school trimming procedure used for the HSTS 2000 sample weights was used for the NAEP 2000-linked weights. Subject-specific school weights were trimmed within NAEP region for public schools and within Catholic/non-Catholic for nonpublic schools. Because NAEP-linked weights are subject specific, trimming was done separately by assessment subject. For a description of the school trimming procedure, see section 6.1.1.3.

The subject-specific school base weight needed in the trimming procedure was calculated as:

$$SCHBWT_s = PSUWGT\_M \times RSCHWT \times SCH\_WT \times HSTSWT \times ASBJWT_s \times STYWT \quad (6.20)$$

where *PSUWGT_M*, *RSCHWT*, *SCH_WT*, and *HSTSWT* were the weighting components as defined in section 6.1.1.1, and $ASBJWT_s$ and *STYWT* were the weighting components defined in section 6.1.2.2.

One school weight was trimmed for each subject. For both subjects, the same public school in the West NAEP region was trimmed in the HSTS 2000 sample weighting process. Each student in the mathematics assessment in this school had his or her weight trimmed by a factor of 0.5169, and in the science assessment by 0.5167.

6.1.2.5 School Nonresponse Adjustment

In this procedure, subject-specific weights of cooperating schools were adjusted to account for eligible schools that did not cooperate in a given subject. A school was considered a cooperating school if it participated in NAEP 2000 for the given subject, participated in the HSTS 2000, and its students maintained the link between the HSTS transcripts and the NAEP assessment scores. It was considered a noncooperating school if it did not participate in either NAEP 2000 (assuming it was assigned the given subject) or the HSTS 2000, or if the link between HSTS and NAEP was missing for students in the school.

The nonresponse procedure was carried out separately by subject. To the extent possible, the definitions of the school nonreponse cells were the same as those definitions used for NAEP 2000 to maintain consistency with the NAEP weights. In public schools, nonresponse cells were defined by the first three digits of PSU stratum (reflecting NAEP region, MSA status, and various socioeconomic characteristics such as percent minority and percent of persons 25 years or older with a college degree) and sample type. In nonpublic schools, the cells were defined by reporting group (Catholic, Lutheran, Conservative Christian, Other Religious, Nonsectarian, and Independent) and sample type.

Occasionally, collapsing of initial cells with adjacent cells was necessary to improve the stability of the adjustment factors. Most cells were collapsed because they contained a small number of cooperating schools, while less often cells with low response rates (and hence large factors) were collapsed. Collapsing procedures were implemented if a nonresponse cell had less than six cooperating schools, or its adjustment factor exceeded 3.0. After collapsing, 25 final nonresponse cells were formed for schools assigned to mathematics and 24 for schools assigned to science.

In each nonresponse cell h, the nonresponse adjustment factor was calculated as follows:

$$SCH\_NRAF_{s,h} = \frac{\sum_{B_{s,h}} SCHBWT_{s,h,i} \times SCH\_TRIM_{s,h,i}}{\sum_{C_{s,h}} SCHBWT_{s,h,i} \times SCH\_TRIM_{s,h,i}} \quad (6.21)$$

where

- $SCHBWT_{s,h,i}$ = the school base weight for school *i* in nonresponse cell *h* for the given subject;
- $SCH\_TRIM_{s,h,i}$ = the school trimming factor for school *i* in nonresponse cell *h* for the given subject;
- $B_{s,h}$ = the set of all eligible schools assigned the given subject in nonresponse cell *h*; and
- $C_{s,h}$ = the set of all eligible schools assigned the given subject in nonresponse cell *h* that ultimately participated.

Nonresponse adjustments assume that nonresponse occurs at random within the categories within which adjustments are made (Little and Rubin 1987). Some degree of bias could result to the extent that this assumption is false. Table 24 shows the distribution of the school nonresponse adjustment factors for the linked samples by subject.

Table 24. Distribution of school nonresponse adjustment factors for the NAEP-linked HSTS samples by subject: 2000

	School nonresponse adjustment factors	
Distribution	Mathematics	Science
Minimum	1.0000	1.0000
25th percentile	1.1246	1.1547
50th (median)	1.3431	1.3159
75th percentile	1.4712	1.4712
Maximum	2.6090	2.7663
Mean	1.4153	1.4060

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.2.6 School Substitution Adjustment

Recall from section 6.1.1.5 that a school substitution adjustment was necessary because the weights of students in a substitute school should reflect what the weights of the students in its

corresponding original school would have been if the original school participated. Of the 12 participating substitute schools in the HSTS 2000 sample, 11 schools were assessed in mathematics while all 12 schools were assessed in science. Every student in a participating substitute school received the appropriate substitution adjustment factor. Table 25 shows the distributions of the substitution adjustment factors for the substitute schools by subject. All students in the original schools received a *SUBADJ* factor of 1.0.

Table 25. Distribution of substitution adjustment factors for the HSTS NAEP-linked samples by subject: 2000

	School nonresponse adjustment factors	
Distribution	Mathematics	Science
Minimum	0.6511	0.6511
25th percentile	0.8417	0.8559
50th (median)	1.0200	1.0242
75th percentile	1.1100	1.1600
Maximum	1.4234	1.4234
Mean	0.9827	1.0175

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.2.7 Year-Round School Weighting Adjustment

This weighting adjustment inflates the weights to account for students in year-round schools who were off track at the time of student sampling. The same two year-round schools that were adjusted for in the HSTS 2000 sample weighting in section 6.1.1.6 appeared in the linked samples for both subjects. Students in these two schools received a year-round weighting factor of 1.4925. The students not in year-round schools received a factor of 1.0.

6.1.2.8 Student Nonresponse Adjustment

This procedure adjusts for "student nonresponse" in the linked HSTS 2000 samples. The response disposition of a student depended on his/her response disposition for both the HSTS 2000 and NAEP 2000. In general, eligible students who participated in the HSTS 2000 and NAEP 2000 were considered respondents for the linked samples. Eligible students who did not participate in both were considered nonrespondents. Students who were ineligible for the HSTS 2000 (i.e., those who did not

graduate) were ineligible for the linked samples regardless of their NAEP 2000 dispositions. Ineligible students were not included in this adjustment but were retained since they were needed for the poststratification process. Table 28 summarizes student response dispositions based on the student dispositions for the HSTS 2000 and NAEP 2000.

Table 26. Student-level response dispositions for the linked samples by HSTS 2000 and NAEP 2000 student dispositions

	NAEP 2000 response disposition		
HSTS 2000 response disposition	Assessed	Absent	Excluded
Graduate with transcripts	Respondent	Nonrespondent	Respondent
Graduate with missing transcripts	Nonrespondent	Nonrespondent	Nonrespondent
Nongraduates	Ineligible	Ineligible	Ineligible

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

The nonresponse procedure was carried out using the same procedure as NAEP to the extent possible to maintain consistency with the NAEP weights. Student nonresponse adjustments were done separately by subject. The initial student nonreponse cells were defined by sample type, school nonresponse cell, race/ethnicity, and age. Race/ethnicity and age were broken down into two categories. The first race/ethnicity category was White or Asian Pacific Islander, and the second category was Black, Hispanic, American Indian, or other. The age category was broken down by students born on or before September 30, 1981 and those students born after that date.

An initial cell was collapsed with an adjacent cell if it had less than 30 responding students as defined above or if its adjustment factor exceeded 2.0.

In each nonresponse adjustment cell h, the student nonresponse adjustment factor $STU\_NRAF_{s,h}$ for subject s was calculated as follows:

$$STU\_NRAF_{s,h} = \frac{\sum_{B_{s,h}} STU\_BWT_{s,h,i} \times SCH\_TRIM_{s,h,i} \times SCH\_NRAF_{s,h,i} \times SUBADJ_{h,i} \times YRRND\_FC_{h,i}}{\sum_{C_{s,h}} STU\_BWT_{s,h,i} \times SCH\_TRIM_{s,h,i} \times SCH\_NRAF_{s,h,i} \times SUBADJ_{h,i} \times YRRND\_FC_{h,i}} \quad (6.22)$$

where

- $B_{s,h}$ = the set of eligible HSTS 2000 linked students (i.e., respondents and nonrespondents) in the linked sample for subject s in student nonresponse cell h;

- $C_{s,h}$ = the set of respondents in the linked sample for subject *s* in student nonresponse cell *h*;
- $STU\_BWT_{s,h,i}$ = the student base weight for the linked sample for subject *s*, as defined in section 6.1.2.2, for student *i* in student nonresponse cell *h*;
- $SCH\_TRIM_{s,h,i}$ = the school trimming factor for subject *s*, as defined in section 6.1.2.4, for student *i* in student nonresponse cell *h*;
- $SCH\_NRAF_{s,h,i}$ = the school nonresponse adjustment factor for subject *s*, as defined in section 6.1.2.5, for student *i* in student nonresponse cell *h*;
- $SUBADJ_{h,i}$ = the school substitution adjustment, as defined in section 6.1.2.6, for student *i* in student nonresponse cell *h*; and
- $YRRND\_FC_{h,i}$ = the year-round school weighting adjustment, as defined in section 6.1.2.7, for student *i* in student nonresponse cell *h*.

The HSTS 2000 linked sample weight used in the student nonresponse procedure reflected the linked student base weight and all school- and student-level weighting adjustments prior to this adjustment.

Each respondent in a nonresponse cell received a nonresponse adjustment factor, as calculated above, corresponding to that cell. Ineligible students received a nonresponse adjustment factor of 1.0.

Table 27 presents the distribution of the student nonresponse adjustment factors for the linked samples. These adjustment factors were calculated for 6,965 HSTS 2000 sampled students who took the NAEP 2000 mathematics assessment and 8,486 HSTS 2000 sampled students who took the NAEP 2000 science assessment.

Table 27. Distribution of student nonresponse adjustments by assessment: 2000

Distribution	Student nonresponse adjustment factors	
	Mathematics	Science
Minimum	1.0113	1.0278
10th percentile	1.0572	1.0672
25th percentile	1.0977	1.1295
50th percentile (median)	1.2209	1.2427
75th percentile	1.3809	1.3856
90th percentile	1.5507	1.5840
Maximum	1.9427	1.8559
Mean	1.2686	1.2851

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.2.9 Student Trimming Adjustment

The student trimming procedure used for the HSTS 2000 sample weights, as described in section 6.1.1.8, was also used for the HSTS NAEP 2000 linked weights. Separately by assessment subject, student weights (through the student nonresponse adjustment) were trimmed if their school contributed more than a specified proportion to the variance on the estimated number of students within a given domain. The domains were NAEP region for public schools and Catholic/non-Catholic for nonpublic schools. School-level estimates of students were calculated by summing the weights of students (i.e., graduates and nongraduates) in the school.

For the mathematics-linked samples, students from two public schools had their weights trimmed. There was one school from the Southeast region where all 55 students received a trimming factor of 0.93780. At another school from the West NAEP region, all 52 students received a trimming factor of 0.93679. For the science-linked samples, there was one public school from the West region that required trimming. All 68 students in this school received a trimming factor of 0.80330.

6.1.2.10 Poststratification Adjustment

The poststratification adjustment procedure described in section 6.1.1.9 was applied to each of the four sets of linked weights separately, using the same poststratification cell definitions, population

control totals,[12] and factor calculations. Table 28 shows the poststratification adjustment factors for each of the four linked HSTS 2000 samples.

Table 28. Poststratification adjustments for the linked samples: 2000

Post-stratification cell	Race/ethnicity	Census region	Age group	Poststratification adjustment factors			
				Mathematics		Science	
				R2 sample	R3 sample	R2 sample	R3 sample
1	Blacks	All	≤ 17	1.2684	1.2397	1.2449	1.2540
2	Hispanics	All	≤ 17	0.9463	0.9496	0.9741	0.9877
3	Other race	All	≤ 17	0.8856	0.8355	0.8087	0.7844
4	Non-Hispanic Whites	Northeast	≤ 17	1.0519	1.0784	1.1172	1.1531
5	Non-Hispanic Whites	Midwest	≤ 17	1.2916	1.2947	1.2413	1.2316
6	Non-Hispanic Whites	South	≤ 17	1.1641	1.1879	1.2001	1.2414
7	Non-Hispanic Whites	West	≤ 17	1.6608	1.6923	1.7195	1.7405

SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.1.2.11 Final HSTS NAEP-Linked Weights

Final HSTS NAEP-linked sampling weights were assigned to all students in the HSTS 2000 study for whom usable transcripts were received and who were assessed (or excluded) in the given NAEP 2000 subject. The weights were computed for each linked sample as follows:

$$FSTUWT = STU\_BWT \times SCH\_TRIM \times SCH\_NRAF \times SUBADJ \times YRRND\_FC \times STU\_NRAF \times STU\_TRIM \times RPTPS\_AD \quad (6.23)$$

where

- *STU_BWT* = NAEP-linked student base weight (as defined in section 6.1.2.2);
- *SCH_TRIM* = School trimming adjustment factor (as defined in section 6.1.2.4);
- *SCH_NRAF* = School nonresponse adjustment factor (as defined in section 6.1.2.5);
- *SUBADJ* = School substitution adjustment factor (as defined in section 6.1.2.6);
- *YRRND_FC* = Year-round school adjustment factor (as defined in section 6.1.2.7);
- *STU_NRAF* = Student nonresponse adjustment factor (as defined in section 6.1.2.8);

[12] See table 19 in section 6.1.1.9 for the control totals for each poststratification cell.

- *STU_TRIM* = Student trimming adjustment factor (as defined in section 6.1.2.9); and
- *RPTPS_AD* = Poststratification factor (as defined in section 6.1.2.10).

The distributions of the final student weights for four HSTS 2000 linked samples are given in table 29. The tables include the count of eligible participating students (i.e., graduates with usable transcripts and who participated in a given NAEP 2000 assessment), the total sum of the weights over all of these students, the minimum and maximum weights, and the quartiles for these weights. The coefficient of variation (CV), computed as the standard deviation of the weights divided by the mean of the weights, is also included.

Table 29. Distribution of final HSTS student weights for the linked samples: 2000

	Mathematics		Science	
Sample distribution	R2 reporting sample	R3 reporting sample	R2 reporting sample	R3 reporting sample
Number of graduates with transcripts and assessed in NAEP	6,368	6,411	7,809	7,788
HSTS linked sample weights				
Total	2,988,332	2,989,001	2,971,480	2,981,026
Minimum	98.8085	87.2417	74.3464	64.2954
25th percentile	259.3462	253.4739	200.9414	202.4979
50th percentile (median)	379.4223	378.1300	304.0154	306.9963
75th percentile	588.3433	590.6721	503.6585	510.1675
Maximum	4397.2223	4480.4415	1912.8578	2782.1919
Mean	469.2733	466.2301	380.5199	382.7716
Coefficient of Variation	64.51	64.93	62.14	63.55

NOTE: The coefficient of variation measures the spread of a set of data as a proportion of its mean. This percentage is the ratio of the sample standard deviation to the sample mean.
SOURCE: U.S. Department of Education, Institute of Education Statistics, National Center for Education Statistics, High School Transcript Study, 2000.

6.2 Variance Estimation

Student estimates based on the HSTS 2000 are subject to sampling error because they are derived from a sample, rather than from the whole population. The variance is a measure of sampling error and, for the most part, determines the reliability of an estimate. Sampling variance indicates how

much a population estimate for a given statistic is likely to change if it had been based on another equivalent sample of individuals drawn in exactly the same manner as the achieved sample.

Since the HSTS 2000 used a complex sample design with several stages of sampling, unequal selection probabilities, and complex weighting procedures, use of standard textbook formulas or standard routines in software packages such as SAS and SPSS generally underestimates the true variance of survey estimates and should not be used. Replicate weights have been provided for each set of sample weights to allow users to compute variances for HSTS 2000 estimates. Use of replicates to calculate variances is generally known as replication. The particular method used for HSTS 2000 was the stratified jackknife assuming two PSUs per stratum (Krewski and Rao 1981), the same method used for the main NAEP 2000.

6.2.1 Jackknife (JK2) Replication Method

The basic idea behind replication is to select subsamples repeatedly from the whole sample, calculate the statistic of interest for each subsample, and then use the variability among the subsample or replicate statistics to estimate the variance of the full sample statistic. Different ways of creating subsamples from the full sample result in different replication methods. The subsamples are called replicates and the statistics calculated from these replicates are called replicate estimates.

The stratified jackknife replication method used for HSTS 2000, known also as the JK2 replication method, assumes that the population of PSUs, the first stage units, is grouped in L variance strata with two PSUs (or variance units) selected from each stratum. In general, a replicate estimate is formed by randomly selecting one variance unit in a variance stratum. The weight of the selected variance unit is doubled, the weight of the nonselected variance unit is multiplied by zero, and the weights for the variance units in the remaining variance strata are not modified. This process is repeated for each variance stratum. If there are L variance strata, then L replicates are created.

The JK2 replication method, as well as any of the other replication methods, is implemented by using replicate weights. Each replicate weight corresponds to a given replicate. The estimated sampling variance of some statistic t is calculated by taking the sum of M squared differences (where M is the number of replicate weights developed):

$$\hat{V}ar(t) = \sum_{i=1}^{M} (t_i - t)^2 \qquad (6.24)$$

where t_i denotes the statistic of interest obtained using the i^{th} set of replicate weights and t denotes the statistic obtained using the set of full sample weights.

6.2.2 Calculating Replicate Weights

Replicate weights for a given HSTS 2000 sample were created by generating random samples of the original sample. In all, 62 replicate weights were created on each student record in an HSTS 2000 data set. Thirty-six replicates were designed to reflect the variance contribution arising from sampling PSUs (generally known as between-PSU variance). The remaining 26 replicates were designed to reflect the variance contribution arising from sampling schools within the 22 certainty PSUs (generally known as within-PSU variance). This variance replication scheme was the same one traditionally used for the national main NAEP 2000 assessment samples.

The creation of the 36 variance strata for the noncertainty PSUs involved pairing noncertainty PSUs in a manner that models a two PSU per stratum design in which PSUs are drawn with replacement. The HSTS 2000 samples used the main NAEP 2000 pairings, where PSUs were paired based on similar stratum characteristics. The 36 pairs of PSUs were formed by putting together PSUs from adjacent strata within NAEP region and metro status. Adjacent strata had similar socioeconomic characteristics such as proportion minority population, population change since 1980, per capita income, civilian unemployment rate, educational attainment, and unemployment rate. Each PSU in a pair was randomly assigned to one of two different variance units (1 or 2). Each PSU pairing was referred to as a variance stratum, and each PSU in a variance stratum was referred to as a variance unit.

The procedure for creating the 26 variance strata for the certainty PSUs was analogous but somewhat more complex. The first stage units in certainty PSUs were schools, and so schools were paired to form variance strata under the JK2 model. For the 22 certainty PSUs in each HSTS 2000 sample, schools were listed in order of selection, and successive schools were paired within certainty PSUs. If there were an odd number of schools within a certainty PSU, the last three schools were grouped into a triple. Each school grouping was referred to as an initial variance stratum. Each school in a pair (or triple) was randomly assigned to 1 of 2 (3) different variance units (1, 2, or (3)). Since the number of initial

variance strata greatly exceeded the desired number of variance strata (26), the initial strata were systematically assigned to 26 "combined" variance strata.[13] To distinguish between the two types of variance components, the 26 variance strata for the certainty PSUs were labeled 1 through 26, and the 36 variance strata for the noncertainty PSUs were labeled 27 through 62.

Replicate student base weights (i = 1–62) for a student assigned to variance stratum with two first-stage sampling units were calculated as below. *STU_BWT* was the student base weight for a given HSTS 2000 sample, as described in section 6.1, which reflected the various stages of selection.

$$STU\_BWT_{rep\_i} = \begin{cases} 0 & \text{if student is in variance unit 1 of variance stratum i} \\ 2 \times STU\_BWT & \text{if student is in variance unit 2 of variance stratum i} \\ STU\_BWT & \text{if the student is not in variance stratum i} \end{cases} \quad (6.25)$$

When a stratum contained three first-stage sampling units, students in the stratum had their weights adjusted for two sets of replicates. Replicate student base weights (i = 1–62) for a student assigned to variance stratum with 3 first-stage units were calculated as follows:

$$STU\_BWT_{rep\_i} = \begin{cases} 0 & \text{if student is in variance unit 1 of variance stratum i} \\ 1.5 \times STU\_BWT & \text{if student is in variance unit 2 or 3 of variance stratum i} \\ STU\_BWT & \text{if the student is not in variance stratum i} \end{cases} \quad (6.26)$$

The final replicate weights for a given HSTS 2000 data set were calculated by applying the same weighting adjustment procedures described in section 6.1 to each set of replicate base weights. By applying the weighting procedures on each set of replicate base weights, variance estimates reflected intended effects of the weighting adjustments.

[13] Initial variance strata comprising 3 schools were assigned 2 variance strata so that 2 replicates are created for each of these strata. This is one common approach to handle 3 PSUs per stratum.

7. GUIDE TO THE DATA FILES AND CODEBOOKS

This chapter describes the content and organization of the 2000 High School Transcript Study (HSTS) data files and codebooks. It also details the process for obtaining the data files.

7.1 Restricted-Use Data Files

All NAEP files, including the NAEP HSTS 2000 data files, are only available to users as restricted-use data files. Restricted-use data files contain school and student variables that cannot be released to the public, but are made available to educational researchers. By federal law, the schools and students that participated in HSTS 2000 are to remain confidential. The HSTS 2000 data files cannot contain any information that directly identifies a school or student, such as school name, school address, or student name.

Because of confidentiality legislation, secondary users who wish to obtain a copy of the restricted-use data files must apply for an NCES restricted data license. If an organization does not already have a restricted data license, it is necessary to obtain a copy of the "Restricted-Use Data Procedures Manual." There is a four-page checklist in this document that details the steps involved in obtaining a license. The manual may be viewed and downloaded from the NCES web site at http://nces.ed.gov/statprog/rudman, or a copy may be requested from the following contact individual:

Cynthia Barton (202) 502-7307
cynthia.barton@ed.gov

If your organization already has a restricted data license, you may need only to have it amended to add datasets or name of individuals as authorized users of the data. Note that, in college or university settings, only faculty can serve as the primary project officer. Graduate students may be listed as authorized users only.

To obtain a restricted data license (or to amend an existing license), a secondary user generally must send a letter addressed to the NCES Data Security Office, formally requesting the data. The mailing address of the Data Security Office is:

NCES Data Security Office
Department of Education/NCES/ODC/SSP
1990 K Street NW
Room 9061
Washington, DC 20006-5574

Please include the following information in your request:

- The name of the data set(s) you want to use;
- The purpose for the loan of the data;
- The length of time you will need the data;
- The computer security plan you will follow;
- The list of authorized users;
- An affidavit of nondisclosure for each person, promising to keep the data completely confidential; and
- If you are amending a license, the license number you want to amend.

7.2 Content and Organization of the Data Files

Data from the 2000 High School Transcript Study were organized into 11 data files encompassing the different levels of information:

- Course Offerings File;
- Master CSSC File;
- School File;
- SD/LEP Questionnaire File;
- Student File;
- Transcript File;
- Tests and Honors File;
- Linked Weights File—Mathematics (R2);

- Linked Weights File—Mathematics (R3);
- Linked Weights File—Science (R2); and
- Linked Weights File—Science (R3).

In addition, there are two NAEP files: NAEP 2000 Mathematics Data File, and NAEP 2000 Science Data File. These files provide information on students' testing participation in the NAEP Mathematics and Science assessments.

Except for the Master CSSC File (which is not related to individual schools or students), all files can be linked by primary sampling unit (PSU) and school identifiers. The SD/LEP Questionnaire, Student, Transcript, Tests and Honors, and Linked Weights Files can be linked by student identifiers; and the Master CSSC File can be linked to either the Course Offerings or the Transcript File by CSSC number.

To identify a specific school, the PSU and school IDs must be used in combination. Each school had a unique PSU/school ID combination. All student IDs were unique. For students in the 248 schools that were fully linked to NAEP, student IDs matched their 10-digit NAEP booklet numbers. All other students were assigned 10-digit IDs beginning with 990.

Weights, developed using the procedures described in chapter 6, are contained in the Student File and the four Linked Weights Files. The final student weight (FINSTUWT) is given in the Student File, and a final usable linked weight (FINLNKWT) is provided in each of the four Linked Weights Files. All HSTS 2000 data analyses can be weighted up to national 12th-grade student totals. The final student weight should be used in analyses involving only high school transcript data. The weights in the Linked Weights File should be used in analyses involving both transcript data and data obtained from the NAEP 2000 data files.

7.3 Course Offerings File

The Course Offerings File contains one record for each course listed in the school's course catalog or appearing on a student's transcript as a nontransfer course taken at that school. Organized by school, each of the file's 68,238 records contains the following information:

- 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),
- the catalog type (whether the catalog is a district-level catalog, a school catalog, or a list of courses generated by the school),
- the location of the course (including various off-campus locations);
- the language of instruction;
- whether it was a remedial or below grade-level course;
- whether it was an honors-level course; and
- whether it was part of an instructional sequence.

The combination course flag (i.e., the course is composed of more than one part, requiring more than one CSSC code for accurate description) was dropped from the HSTS 2000 Course Offerings file. The flag was of no value since courses could be properly assigned to a single CSSC code. The file is sorted by the PSU and school IDs.

The Course Offerings File is a complete listing of courses offered in all participating schools that provided school-level course catalogs, as well as any nontransfer courses listed on the transcripts not otherwise appearing in the catalogs. It should be noted that schools may not offer all courses that are in their catalogs. For example, in a high school with grades 10 through 12 whose students all take 9th grade courses in junior high, the 9th-grade courses were not treated as transfer courses, but appeared as if they were offered by the high school. This treatment provided a more balanced picture of the courses available to students in 4 years of high school than would be provided by treating such courses as transfer courses. For the 31 schools from which no catalog was received, the list of unique course titles appearing on the sampled transcripts was the only available source of course offering entries.

7.4 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. The CSSC is a modification of the Classification of Instructional Programs (CIP) that is used for classifying college courses. Each course that appears on a student transcript is assigned a unique six-digit code based on the course content and level. Course catalogs and other materials from the participating schools are used to determine the content and level of courses at each school. The system is adaptable so that new or revised courses are easily incorporated. There are 2,268 records in the file, sorted by CSSC number. In addition to the original 6-digit CSSC codes created in 1982, the file contains the codes added for the 1987, 1990, 1994 and 1998 studies. Appendix C contains more information about the CSSC codes.

Two new CSSC codes were added for the HSTS 2000, one in science and one in computer-related studies. These codes were added when courses were encountered on the transcripts that were clearly different from codes already contained in the master CSSC list. Since there were no new subject areas identified in the HSTS 2000, no new 2-digit or 4-digit categories were added. Five duplicate and unused codes were dropped in 2000.

The Master CSSC File is organized by the CSSC code and contains four variables: the CSSC course code, the special education flag, the standard course title, and the sequence flag. The 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, and the special education flag was set to 0 or 2.[14] Any course not so limited had the special education flag set to 1.

Consistent with the 1990, 1994 and 1998 transcript studies, all CSSC entries in the HSTS 2000 were coded with a sequence flag. A zero value for the sequence flag indicated that the course was not part of an instructional sequence. A one indicated that the course was the first course in an instructional sequence, and a 2 indicated that the course as an advanced course in an instructional sequence (i.e., not the initial course in the sequence).

[14] 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.

7.5 School File

The School File contains one record for each of the 277 participating schools. Sorted by PSU and school ID, the file includes school variables gathered during the transcript study, as well as the school's responses to the NAEP School Questionnaire. Schools that did not participate in NAEP were also asked to fill in the NAEP School Questionnaire. A copy of the questionnaire is in appendix B.

Because the HSTS 2000 school sample was a subsample of the original NAEP 2000, each HSTS 2000 participating school had assigned a three-digit Primary Sampling Unit ID and a three-digit NAEP School ID. To form the HSTS School ID, an additional digit was added to the NAEP School ID to indicate the school type (public or nonpublic) and sampling status. When concatenated together, each HSTS school had a unique seven-digit ID.

Changes were made in the composition of the School File for HSTS 2000. In previous studies, the Quality Education Data[15] national school database supplied information on the number of teachers (NUMTEACH) and student enrollment (ENROLL) variables. For the HSTS 2000, the 1997-98 Common Core of Data (CCD) public school database and the 1997-1998 Private School Universe Survey provided data for these variables. There were also some changes made in the NAEP School Questionnaires that are reflected in the School files of various transcript studies, including HSTS 2000.

7.6 SD/LEP Questionnaire File

School special education staff members were asked to fill out an SD/LEP Questionnaire for each disabled student and each student with limited English proficiency sampled for NAEP. In addition, there were seven students who were not linked to NAEP that had SD/LEP information. The SD/LEP Questionnaire File contains one record each for 2,561 students. The file is sorted by PSU, school, and student ID, and contains data from the completed questionnaires.

[15] 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. The listing 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.

7.7 Student File

The Student File contains one record for each of the 23,522 high school 12th-grade students who were identified for the HSTS 2000. The file is sorted by the PSU, School ID, and Student ID variables. Each record in the file contains demographic information, sampling information, student weights, and replicate weights for variance estimation. The file also contains a series of derived variables, including summaries of the student's coursetaking record by major educational topic, as taken from the student's transcript data. Because 2,513 transcripts were not received, only 21,009 students have full transcript information on their student records.

In NAEP-linked schools (i.e., schools that participated in and retained their link to the NAEP 2000 assessment), each student received a unique 10-digit booklet ID. This booklet ID became the HSTS 2000 Student ID. A different procedure was required for schools that either did not retain their materials that linked selected students to their specific NAEP IDs or did not participate in NAEP 2000. In those schools, a new sample of students was selected. Those students were assigned ID numbers that started with 9900000001 and continued sequentially to 9900000800. While this procedure ensured unique student ID numbers, for purposes of identifying HSTS 2000 students across all files, students are identified by the combination of PSU, school ID, and student ID variables.

The file also contains a flag indicating whether or not the student was disabled (HCFLAG) and a condition variable indicating the specific nature of the disability when applicable (HCTYPE).[16] The Student Disability Description variable (HCTYPE) in the HSTS 2000 Student File contains different categories than the similar variable in the HSTS 1998 file. Because no category was defined for "Not Disabled," the format of the HSTS 2000 Student Disability Description variable completely differs from the previous format. In addition, new disability categories exist in the 2000 file, namely categories 09 (Autism), 10 (Developmental Delay), 11 (Other Health Impaired), and 88 (Not Reported).

The weights included on the Student File are for all HSTS 2000 students, covering both students that could be linked to the NAEP assessment and those students that could not be linked. Analyses of 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

[16] The values of the disabling condition code are 00 = Multidisabled, 01 = Learning Disabled, 02 = Hearing Impaired, 03 = Visual Impaired, 04 = Speech Impaired, 05 = Mental Retardation, 06 = Emotional Disturbance, 07 = Orthopedic Impaired, 08 = Traumatic Brain Injury, 09 = Autism, 10 = Developmental Delay, 11 = Other Health Impaired, 12 = Other, 88 = Not Reported, 99 = Not Collected.

contained in the appropriate Linked Weights Files. The final student weight for each student is the variable FINSTUWT, while the replicate weights used for variance estimation are REPWT1 – REPWT62. Note that 231 students for whom no transcripts were obtained had final student weights (FINSTUWT) of zero. Only transcripts fully documenting at least 3 years of high school received positive weights.

7.8 Transcript File

The Transcript File contains one record for each course appearing on the sampled students' transcripts. It is an extremely large file, containing 995,035 records. Courses are uniquely identified by PSU, school, student ID, and course sequence number (within students), and the Transcript File records are sorted by those variables. Each course record includes the following variables:

- grade level when taken,
- school year when taken,
- course title,
- grade received (original and standardized),
- credit received (original and standardized),
- course CSSC code,
- whether the course was taught off campus,
- whether the course was taught in a language other than English,
- whether the course was remedial or below grade level, and
- whether the course was an honors course.

7.9 Test and Honors File

The Test and Honors File contains information on standardized test scores and honors that appear on high school transcripts. Of the transcripts collected, 5,587 (23.75%) transcripts contained either

standardized test scores or notations regarding honors and awards that students received. Because of the relatively small percentage of transcripts represented, the data in this file should be used with caution.

As in the Student File, students in the Test and Honors File are identified by the combination of PSU, school, and student ID variables. The file is sorted by these identifier variables. Each test or honor entry on a transcript is identified with a unique sequence number. The combination of PSU, school, student ID, and test/honor sequence number allows for a unique ID for each test or honor within the file. Entries are sorted by sequence number within student. Each entry also contains an indicator of the record type ("T" = test, "H" = honor), the month and year of the test or honor (if available), and a 40-character description of the honor or the test.

Test scores were provided for most tests. It was not always possible to give meaningful entries for some test scores (e.g., some schools reported Standard Reading and Writing Assessment (SRA) tests with percentiles while other schools reported scaled scores). The subtests which are reported also varied tremendously. Nevertheless, complete scores are provided for the Preliminary Scholastic Aptitude Test (PSAT) math and verbal subtests, the Scholastic Aptitude Test (SAT) math and verbal subtests, and the American College Test (ACT) composite 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 19,381 records (18,371 standardized tests and 1,010 honors).

7.10 Linked Weights Files

The four Linked Weights Files contain the set of weights needed to perform analyses on the subset of HSTS 2000 students fully linked to the NAEP 2000 assessment. Because different sets of schools were eligible to participate in the NAEP and HSTS studies, and because different sets of schools chose to participate in each study, different school-level nonresponse adjustments were used when constructing student weights. For similar reasons, different student-level nonresponse adjustments need to be used when constructing student weights. Furthermore, since the national main NAEP 2000 study consisted of two parallel sets of assessments (Mathematics and Science), separate sets of weights need to be used for each assessment. Within these two assessments, SD/LEP students in the sample are handled two ways—included students and excluded students—based upon whether or not accommodations was provided. A separate set of weights is provided for students based upon their inclusion status from the NAEP assessments on the basis of a disability or limited English proficiency.

The linked weights are to be used whenever the analysis uses NAEP data. The linked weights are created to analyze each NAEP subject separately, with or without accommodations. On the other hand, the nonlinked weights on the Student File are used when analyzing transcript data only (i.e., without regard to NAEP data). One difference between the processes for creating linked and nonlinked weights is in the treatment of nonresponse. The linked weights are adjusted to account for nonrespondents, where nonrespondents are eligible students with incomplete transcripts or eligible students that were absent in NAEP. The nonlinked weights are adjusted to account for nonrespondents, where nonrespondents are eligible students with incomplete transcripts.

When combined, the four Linked Weights Files contain one record for each of the 22,010 graduates who had NAEP booklet numbers. As in the Student File, students are identified by the combination of PSU, school, and student ID variables. The files are sorted by these identifier variables. The first digit of the student ID identifies the assessment in which the student participated. A value of 1 indicates a mathematics assessment, whereas a value of 2 indicates a science assessment.[17] For ease of use, these files also contain the demographic variables included on the Student File. The final usable linked weight variable is FINLNKWT.

7.11 NAEP Mathematics and Science Data Files

There are two NAEP data files— the 2000 NAEP Mathematics Data File and the 2000 NAEP Science Data File—which contain proficiency scores for each year 2000 graduate who participated in a NAEP assessment in a school that was fully linked to the HSTS 2000. The NAEP proficiency scores, otherwise known as plausible values, are not merged directly from the NAEP files to the transcript files. The scores are first adapted to the transcript data prior to merging them to the transcript files.

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.[18] The conditioning process has the effect of increasing the bias when analyses are made between NAEP scores and variables not in the conditioning set. To make the

[17] 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 nonlinked students—that is, students in schools for whom a sample was drawn in the field for the transcript study.

[18] See the forthcoming online NAEP 2000 Technical Report for a detailed discussion of conditioning.

transcript data as usable as possible, a number of transcript study variables were selected to be used in 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 C.
- STUB2001 - STUB 2005 — New Basics Curriculum categories. These variables represent variants of academically oriented course-taking patterns recommended in the *Nation at Risk* report. They are defined in detail in appendix C.

All the variables normally used for conditioning of the NAEP scores were also considered in this conditioning process. Thus, all the variables in the transcript study's Student File can be safely used in analyses involving NAEP scores. Because the variables listed above were also included in the conditioning of NAEP scores for the transcript study, the NAEP scores reported in the HSTS files are slightly different from those scores contained in the records for the same students distributed solely as NAEP data. The overall national scores from the two studies are nearly identical.

As discussed in chapter 3, since 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, rather than the weights contained in the Student File, should be used in analyses comparing the NAEP data to the transcript data. Note that if a complete transcript for a student was not available, his or her weight was 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, be aware of the following differences in naming conventions as shown in table 30.

Table 30. HSTS and NAEP variable naming conventions: 2000

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	4[1]	SCH	3
STUDENT	10	BOOK	3
		BKSER	6
		CHKDIG	1

[1] The School ID is a 3-digit ID to which a fourth control digit is added. It represents the type of school (public vs. nonpublic) and sampling status (original vs. substitute). For the purpose of uniquely identifying a school and matching it to the NAEP School ID, this digit can be dropped. The values of the fourth control digit are 0=Public Original, 1=Catholic Original, 2=Other Nonpublic Original, 3=Public Substitute, 4=Catholic Substitute, and 5=Other Nonpublic Substitute.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

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).

Table 31 summarizes the number of records in each NAEP data file and the corresponding number of nonzero weights in the Linked Weights Files.

Table 31. Comparison of records and nonzero weights in the HSTS Linked Weights Files: 2000

NAEP data file	Number of records	Number of nonzero weights
Mathematics R2	8,941	7,250
Science R2	11,120	8,919
Mathematics R3	8,998	7,294
Science R3	11,136	8,887

NOTE: The R2 reporting sample contains non-accommodated students, while the R3 reporting sample consists of accommodated students.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Transcript Study, 2000.

7.12 Research Issues

This section discusses issues that concern researchers who wish to conduct their own analyses of High School Transcript Study (HSTS) data. In conjunction with the NAEP studies, the transcript studies provide snapshots of high school students' coursetaking patterns from the 1980s through 2000. The HSTS data files contain a wealth of education-based information for researchers to use to inform issues related to coursetaking, access to courses, and achievement. Issues addressed in this section reflect various stages of the research process, including NAEP and HSTS data background information, comparability among previous HSTS studies, recommended areas of data analysis, and approaches and procedures to use in conducting research.

7.12.1 Background Data Collected via NAEP

The design of the NAEP studies does not allow it to report on the performance of individual students. It rather assesses specific populations of in-school students or subgroups of these students, reporting on their performance in selected academic areas. The NAEP results are derived from samples of the study's student populations of interest. NAEP data include students from both public and nonpublic schools. NAEP policy states that, if any doubt exists about a student's ability to participate, the student should be included in the assessment. Beginning with the 1996 assessments, NAEP has allowed more accommodations for SD/LEP students.

In addition to assessing achievement and cognitive abilities in various subject areas, NAEP collects information from selected students, teachers, and principals on a variety of contextual background variables. These variables describe student, teacher, and school characteristics, as well as instructional practices and curricula. When developing the questionnaires used to gather this information, NAEP ensures that the questions do not infringe on respondents' privacy, that they are grounded in research, and that the answers can provide information relevant to the subject being assessed.

Four NAEP questionnaires provide the contextual background variables:

- student questionnaires (with background items and subject specific items);
- teacher questionnaires;

- school questionnaires; and
- students with disabilities/limited English proficiency (SD/LEP) questionnaires.

These questionnaires appear in one of two formats. The student questionnaires appear as separately timed blocks of questions in the assessment booklets. The teacher, school, and SD/LEP questionnaires are printed as separate booklets.

The student questionnaire asks students about their demographic characteristics and experience in the NAEP assessment subjects. Demographic characteristics include race/ethnicity, parental education level, educational resources in the home, and school attendance. Assessment subject experience variables include likes and dislikes, how the subject is studied, and advanced coursetaking.

The NAEP teacher questionnaire gathers background information on teachers and their relevant classroom practices. Teachers are asked about their educational background, training related to the NAEP assessment subjects, and the styles they use to teach those subjects.

The school questionnaire asks about school characteristics. Demographic information collected by the school questionnaire includes the length of the school day and year, school enrollment, absenteeism, dropout rates, size and composition of the teaching staff, tracking policies, curricula, testing practices, and school-wide programs and problems. This questionnaire also collects information about the availability of resources, policies for parental involvement, and participation in federal programs. The principal or another school official completes this questionnaire.

The SD/LEP questionnaires are completed for students by their teachers. For HSTS 2000, questionnaires were completed for SD/LEP students who were selected for participation in NAEP 2000, as well as those students from schools that did not participate in NAEP. SD/LEP students included individuals classified as students with disabilities (SD), limited English proficiency (LEP), educated under Individual Education Plans (IEPs) or with an equivalent classification. The SD/LEP questionnaire gathered information about the student's disability classification. For a student classified as SD, the questionnaire requested information about the student's functional grade level, mainstreaming, and special education programs. For a student classified as LEP, the questionnaire asked about the student's native language, time spent in special education and language programs, and his/her level of English language proficiency.

7.12.2 HSTS Analysis Overview

The HSTS sample comes from a subsample of schools and students included in the corresponding NAEP sample. This subsample allows the linking of NAEP and HSTS data for schools that participated in both studies. Note that not all HSTS school and student records have corresponding NAEP data. Some schools did not participate in the NAEP assessments but yet took part in the transcript study. Other schools participated in the NAEP assessments, but the materials used to link a student to his/her specific NAEP ID were lost.

The HSTS are student-based studies. Weights included on both the student and linked weights files reflect national student totals. Linked weights were created to provide the national student totals of HSTS 2000 students who have NAEP scores, while the student weights provide the national student totals of all students who had transcripts, regardless of whether they had participated in the NAEP. **Although the HSTS includes school and state information, these data must not be used for either school-level or state-level analysis.** Combining the weights of student records within a school or state will not accurately reflect the number of high school graduates from that school or state. Analyses may be conducted with student data at the Census region level, as the sum of weights reflects the number of high school graduates within those Census regions.

Although the HSTS 2000 is a component of the NAEP 2000, most of the data from these studies are maintained and provided as separate studies, while some of the data from the two studies are shared. The NAEP 2000 school and SD/LEP questionnaire data are also included with HSTS 2000 school and SD/LEP files. Data from the NAEP student and teacher questionnaires, however, are not in the HSTS data files. If a researcher has access to both HSTS and NAEP data files, then the contextual background variables on the NAEP student and teacher questionnaires can be linked to the HSTS students. The HSTS data files use the same identification codes to identify schools and students as do the NAEP assessment files, making linking between the two sets of files possible.

7.12.3 HSTS 2000 Tabulations

The forthcoming HSTS 2000 tabulations, *The 2000 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 2000, 1998, 1994, 1990, 1987, and 1982 High School Graduates*, provide a detailed description of the coursetaking patterns of high

school graduates in 2000. The tables also provide, where possible, comparable details from the 1982 High School and Beyond (HS&B) study and the HSTS studies in 1987, 1990, 1994, and 1998, showing changes that have taken place in graduates' coursetaking patterns since 1982. Additional data tables indicate the relationship between coursetaking patterns and student achievement in mathematics and science, the subjects assessed in NAEP 2000.

For the HSTS 2000, an attempt was made to collect high school transcripts from the 23,440 sampled students expected to graduate from high school in 2000. The HSTS 2000 tabulations represented students with complete transcripts. Students whose transcripts did not include course-by-course data for at least 3 full years of high school were excluded. To be consistent with other published analyses, the following rules were adopted for including and excluding students in the analyses that produced the tables:

1. Both public and nonpublic school students were included.

2. Students with special education diplomas, certificates of attendance, and certificates of completion were excluded. Certificates of completion indicate that a student completed the necessary school requirements for graduation, but failed to successfully complete a required state graduation exam.

3. Students with disabilities who received regular or honors diplomas (i.e., those who were not screened out by rule 2) were included.

4. Students with fewer than 16 Carnegie Units were excluded. A Carnegie Unit was a factor used to standardize all credits indicated on transcripts across the study. The Carnegie Unit is defined as the number of credits a student received for a course taken every day, one period per day, for a full school year.

5. Students with zero English credits were excluded.

The HS&B 1982, HSTS 1987, and HSTS 1990 studies initially excluded students who earned more than 32 Carnegie units. The reason given for this exclusion was that the schools these students attended must have shorter class periods than normal schools and use of their data would inflate the estimates. An examination of such schools in the HSTS 1994 study found this reasoning to be incorrect. A majority of these schools were religious private schools, which required stringent graduation requirements and larger courseloads. Starting with the 1994 HSTS study, students with more than 32 Carnegie units were not excluded. Additionally, this exclusion criterion was dropped when the HS&B 1982, HSTS 1987, and HSTS 1990 studies were recoded as part of the HSTS 1994 study.

As previously stated, students whose transcripts did not include course-by-course data for at least 3 full years of high school were excluded from the tabulations. In a few cases, it was determined that a student had not actually graduated and the Exit Status was revised accordingly. It was also found that some students had earned more credits than were required to graduate. These students often had spent substantial amounts of time in both foreign and American high schools. While they were awarded credit for the foreign courses, they were still required to take an essentially American curriculum in order to obtain the American diploma.

Among students with transfer courses, it was determined that, although a student had fewer credits than were required to graduate, the transcript had all the other attributes of a graduated senior. These attributes included student exit status, graduation date, grade point average, and class standing. Credits from transfer schools may not have been recorded on the transcript, or the transferred credits have a different credit assignment than the school of graduation. In these cases, if a careful review of the transcript and the data files showed no data entry or coding errors, and the lack of credits resulted from missing or improperly converted Carnegie credits for the transfer courses, the record was updated. An additional transcript record with undifferentiated credit was added, or the existing transfer credit records were modified to assign the actual number of credits the student had taken.

In summary, for a transcript to be included in the tabulations, it had to meet three requirements: (1) the student graduated with either a Standard Diploma or an Honors Diploma, (2) the student's transcript contained 16 or more Carnegie credits worth of courses, and (3) the student's transcript contained more than 0 Carnegie credits worth of English courses. These tabulations restrictions reduced the number of 2000 graduates represented in the tables to 20,272. These students attended 276 of the 277 schools that had previously been sampled for NAEP 2000.

7.12.4 Areas of HSTS Data Analysis

Data collected by the HSTS offers researchers a unique glance into student coursetaking patterns from one study year to the next. Before analyzing the HSTS data, however, researchers should check whether or not the analysis has already been performed. Many analyses and comparisons have been conducted and can be found in the following reports. The two HSTS 2000 publications listed below contain common educational-related data analyses performed on all six HSTS data sets. It should be noted that, although the NELS:88 transcript data have not been included in the NAEP HSTS reports, they may be used by researchers for the 1992 data point.

The first publication, *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990–2000*, is a printed report available from the National Center for Education Statistics via the web site (http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2004455) and EDPUBS. This report analyzes the changes in course credits earned and grade point averages achieved by high school graduates from HSTS 1990 to HSTS 2000. It also looks at correlation values between the NAEP 2000 mathematics and science assessment scores with various student coursetaking variables.

The second publication, *The 2000 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 2000, 1998, 1994, 1990, 1987, and 1982 High School Graduates*, is an upcoming online publication to be available on the NCES NAEP web site. It details the number of credits earned by high school graduates in various school subject fields and by various school and student characteristics, including gender, race/ethnicity, academic track, type of locale, school type (public/nonpublic), and region of the country. It also contains tables covering graduation requirements, grade point averages, and NAEP 2000 mathematics and science assessment scores.

Both publications focused on high school graduates for their data analyses. To maximize the probability that the data analyses included only detailed high school graduates' transcripts, restrictions were placed on which HSTS student records were included. Only students who graduated with a standard or honors diploma, earned at least 16 Carnegie credits of courses, and earned more than zero Carnegie credits in English courses were used for the analyses. (Note that the HS&B 1982 student data did not contain a graduation status field, so the diploma restrictions were not used for the analyses.) Not all HSTS analyses, however, need to restrict the HSTS data accordingly. Researchers who wish to look at all high school seniors' results, including those students who did not graduate, can run their analyses using all HSTS student records.

The HSTS 2000 data sets offer new possibilities for data analyses that previous HSTS data sets could not offer. One can examine course credits earned and grade point average of high school graduates as defined by two measures of poverty, free/reduced school lunch status and Chapter 1 Title 1 status. Researchers can analyze relationships between the mean NAEP mathematics and science assessment scores by whether or not students took selected mathematics or science courses. Incorporating the HSTS 2000 data sets with the previous HSTS data sets, researchers can track courses by grade level across the transcript studies to determine whether course curricula have changed in the past two decades. Linking the HSTS files with the corresponding NAEP student questionnaires provides new educational-

related variables for data analysis, including parents' education levels, computer usage at home and school, and time spent on homework.

7.12.5 HSTS 2000 Research Approaches and Procedures

Course Codes. High school courses across the country vary by content and level, even when the course title is similar. Therefore, to compare transcripts from different schools and to ensure that each course is uniquely identified, a common course coding system, the Classification of Secondary School Courses, was employed. The CSSC is a modification of the Classification of Instructional Programs (CIP) that is used for classifying college courses. Each course that appears on a student transcript is assigned a unique six-digit code based on the course content and level. The CSSC contains 2,268 course codes. The system is adaptable so that new or revised courses are easily incorporated. Course catalogs and other materials from the participating schools are used to determine the content and level of courses at each school.

For analysis and data presentation purposes, the CSSC uses an outline similar to the Secondary School Taxonomy (SST) to group courses into larger and more useful categories called stubs, such as the English, mathematics, and science stubs. More detailed stubs are sometimes used, for example, to report findings on advanced courses such as Advanced Placement (AP) and International Baccalaureate (IB) courses. As with the previous HSTS data sets, the HSTS 2000 student data file contains the number of Carnegie credits earned by the student in each stub.

All of the courses in each of the transcript studies were coded using the CSSC. Therefore, the coursetaking patterns of the 1982, 1987, 1990, 1994, 1998, and 2000 high school graduates, as measured by the HS&B and HSTS studies, can be compared across years. Appendix C lists all of the CSSC codes used in the high school transcript studies.[19] The final table of the forthcoming companion online report *The 2000 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 2000, 1998, 1994, 1990, 1987, and 1982 High School Graduates* provides the number and percentage of high school graduates who took courses defined by each CSSC code for all six HSTS studies.

[19] The 1992 Second Follow-Up to the National Educational Longitudinal Study also used the CSSC to codes its courses.

NAEP Scale Scores. Because of the design of the NAEP assessments, each student typically responds to only a few questions within any content area, and not all students are asked the same questions. Unlike many traditional assessments, there is no linear transformation between correct/incorrect items and a single score. Using a single student-level score, thus, would result in misleading estimates of population characteristics. Instead, NAEP constructs sets of plausible values (in sets of five) designed to represent the distribution of performance in the population for each subject assessed. A plausible value is a representative value from the potential scale scores for all students in the population with similar characteristics and identical patterns of item response. The NAEP scale scores are further conditioned with student transcript characteristics when used with the High School Transcript Study. NAEP scale scores associated with the HSTS 2000 data, thus, slightly differ from NAEP scale scores associated with NAEP 2000 student data.

Since the statistics describing the performance on the NAEP mathematics and science scales are based on the plausible values, the statistical software used to conduct these analyses must properly compute the statistics for the plausible values.

More information about NAEP 2000, including scale scores, plausible values, and jackknife variance replication can be found in the forthcoming online NAEP 2000 technical report.

Weights. The HSTS 2000 weights are based on the NAEP weights. Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable level of uncertainty for any student information that can be observed without error. There are five sets of weights associated with the HSTS 2000 student data:

- Student weights that encompass all HSTS 2000 students;
- Student weights that include those HSTS students linked to the NAEP 2000 mathematics assessment with SD/LEP accommodations;
- Student weights that include those HSTS students linked to the NAEP 2000 mathematics assessment without SD/LEP accommodations;
- Student weights that include those HSTS students linked to the NAEP 2000 science assessment with SD/LEP accommodations; and

- Student weights that include those HSTS students linked to the NAEP 2000 science assessment without SD/LEP accommodations.

Chapter 6 contains a more detailed description of the weighting procedures.

Analyses conducted on the HSTS 2000 transcript data should use the student weights found on the HSTS 2000 student file (FINSTUWT). When the analyses involve the NAEP scale scores, the linked weights (FINLNKWT on each linked weights file), rather than the student weights, should be used. The linked weights provide the national estimates of high school seniors based on the NAEP mathematics and science assessment samples, while the student weights provide national estimates of high school seniors based on the HSTS student sample. All NAEP assessment tabulations that appear in the upcoming online publication *The 2000 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 2000, 1998, 1994, 1990, 1987, and 1982 High School Graduates* use the linked weights without SD/LEP accommodations. Previous HSTS studies also used linked weights without SD/LEP accommodations for their NAEP assessment research. The NAEP design starting in 2002 phased out the assessment of students without accommodations. Future HSTS studies will follow the new NAEP design which will use only the linked weights with accommodations.

Regardless of whether an HSTS 2000 analysis uses student weights or linked weights, providing the appropriate standard errors requires processing the jackknife replicate weights. Each set of weights includes 62 replicate weights (REPWT1–REPWT62 on the HSTS 2000 student file, LREPWT1–LREPWT62 on the linked weights files).

Statistical Software. Specialized software is required to produce the appropriate statistics from the HSTS 2000 data due to the complex sample design reflected in the jackknife replicate weights and the plausible values of the NAEP scale scores. Standard SAS and SPSS code can produce accurate point estimates but cannot easily produce correct standard errors.

The International Association for the Evaluation of Educational Achievement (IEA) and Australian Council for Educational Research (ACER) developed SAS macros to work with similar kinds of jackknife replicate weight data sets found in international educational assessments like Third International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), and Programme for International Student Assessment (PISA). The IEA software needs to be modified for use with the HSTS 2000 data files. The programs can be downloaded from the IEA web site

(http://timss.bc.edu/TIMSS1/database/UG3.pdf). NCES has also developed software, ***AM Software***, for use on NAEP and NELS:88 (http://am.air.org) that may be applicable to HSTS data in the future.

Commercial software such as ***WesVar*** can also be used for analyzing the HSTS data (http://www.westat.com/wesvar/). Other commercially-available software includes SUDAAN (http://www.rti.org/sudaan/) and STATA (http://www.stata.com/).

8. REFERENCES

Cochran, W.G. (1977). *Sampling techniques*, 3rd ed. New York: Wiley.

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

Gifford, A., Hoachlander, E., and Tuma, J. (1994). *The Secondary School Taxonomy Final Report.* U.S. Department of Education, National Assessment of Vocational Education. Final Report to Congress, Volume II, Chapter 1.

Ingels, S., Dowd, K., L., Taylor, J., Bartot, V., Frankel, M., Pulliam, P. (1995). *National Education Longitudinal Study of 1988, Second Follow-Up: Transcript Component Data File User's Manual* (NCES 95–377). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Jones, C., Knight, S., Butz, M. Crawford, I., and Stephenson, B. (1983). *High School and Beyond Transcripts Survey (1982): Data File User's Manual*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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

Krenzke, T., Rust, K., Wallace, L., and Yansanah, I. (2000). *1998 NAEP sampling and weighting report.* Rockville, MD: Westat.

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

Lee, E.S., Forthofer, R. N., and Lorimer, R.J. (1989). *Analyzing Complex Survey Data*. Newbury Park, CA: Sage Publications.

Legum, S., Caldwell, N., Goksel, H., Haynes, J., Hynson, C., Rust, K., Blecher, N. (1993a). *The 1990 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1990, 1987, and 1982 High School Graduates* (NCES 93–423). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Legum, S., Caldwell, N., Goksel, H., Haynes, J., Hynson, C., Rust, K., and Blecher, N. (1993b). *The 1990 High School Transcript Study Technical Report* (NCES 93–117). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Legum, S., Caldwell, N., Goksel, H., Haynes, J., Hynson, C., Rust, K., and Blecher, N. (1993c). *USER'S MANUAL: 1990 High School Transcript Study* (NCES 93–416). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Legum, S., Caldwell, N., Davis, B., Haynes, J., Hill, T., Litavecz, S., Rizzo, L., Rust, K., and Vo, N. (1997). *The 1994 High School Transcript Study Technical Report* (NCES 97–262). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Legum, S., Caldwell, N., Davis, B., Haynes, J., Hill, T., Litavecz, S., Rizzo, L., Rust, K., and Vo, N. (1998). *The 1994 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1994, 1990, 1987, and 1982 High School Graduates* (NCES 98–532). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Little, R. J. A. and Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley.

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, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

National Assessment of Vocational Education, *Final Report to Congress,* Volume II, chapter 1. Washington, DC: 1994.

National Commission on Excellence in Education. (1983). *A Nation at Risk: the Imperative for Educational Reform*. Washington, DC: U.S. Government Printing Office.

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.

Perkins, R., Kleiner, B., Roey, S., and Brown, J. (2004). *The High School Transcript Study: A Decade of Change in Curricula and Achievement*, *1990–2000* (NCES 2004–455). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Perkins, R., Roey, S., and Brown, J. (2005). *The 2000 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 2000, 1998, 1994, 1990, 1987, and 1982 High School Graduates* (upcoming NCES 2005–482). U.S. Department of Education, National Center for Education Statistics.

Roey, S., Caldwell, N., Rust, K., Blumstein, E., Krenzke, T., Legum, S., Kuhn, J., Waksberg, M., Haynes, J. (2001a). *The 1998 High School Transcript Study User's Guide and Technical Report* (NCES 2001–477). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Roey, S., Caldwell, N., Rust, K., Blumstein, E., Krenzke, T., Legum, S., Kuhn, J., Waksberg, M., Haynes, J. (2001b). *The 1998 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1998, 1994, 1990, 1987, and 1982 High School Graduates* (NCES 2001–498). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Thorne, J., Project Director (1988). *High School Transcript Analysis: The 1987 Graduates*. U.S. Department of Education. Washington, DC: National Center for Education Statistics.

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.

Thurgood, L., Walter, E., Carter, G., Henn, S., Huang, G., Nooter, D., Smith, W., Cash, R.W., and Salvucci, S. (2003). *NCES Handbook of Survey Methods* (NCES 2003–603). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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