High School Transcript Studies (HSTS)

Website: https://nces.ed.gov/surveys/hst/, https://nces.ed.gov/nationsreportcard/hsts/
Updated: October 2017

The value of school transcripts as objective, reliable measures of crucial aspects of students' educational experiences is widely recognized. NCES high school transcript studies collect information contained in the student high school record—i.e., courses taken while attending secondary school; information on credits earned; year and term a specific course was taken; and final grades. When available, information on class rank and standardized scores is also collected. Once collected, this information is transcribed and standardized (e.g., credits and credit hours converted to a common metric) and can be linked back to the student questionnaires or assessment data, although it is generally released to the public at a higher level of aggregation, such as for the nation as a whole, or by student groups.

Transcripts include information that is considered to be the official and fixed record regarding student course-taking behaviors. This information is considered to be more accurate than self-reported student information and represents a record of courses taken by the student. This information can be used to examine course-taking patterns of students and to predict future education outcomes.

Since 1982, NCES has conducted 11 high school transcript studies: seven associated with the National Assessment of Educational Progress (NAEP) and four conducted as part of the Secondary Longitudinal Studies Program. Thus, this chapter is divided into two sections—the first, on the NAEP HSTS; the second, on the transcript studies conducted as part of the Secondary Longitudinal Studies Program.

NAEP HIGH SCHOOL TRANSCRIPT STUDIES

- National Assessment of Educational Progress High School Transcript Studies

SECONDARY LONGITUDINAL STUDIES PROGRAM HIGH SCHOOL TRANSCRIPT STUDIES

- High School and Beyond Longitudinal Study High School Transcript Study
- National Education Longitudinal Study of 1988 High School Transcript Study
- Education Longitudinal Study of 2002 High School Transcript Study
- High School Longitudinal Study of 2009 High School Transcript Study

NAEP High School Transcript Studies

1. OVERVIEW

Since 1982, NCES has conducted seven high school transcript studies (HSTS) associated with the National Assessment of Educational Progress (NAEP). NAEP has collected transcript data in 1987, 1990, 1994, 1998, 2000, 2005, and 2009.

Purpose
Transcripts include information that is considered to be the official and fixed record regarding student course taking behavior. It is considered to be more accurate than student self-report information and represents a record of courses taken by the student. This information can be used to examine course-taking patterns of students and to predict future education outcomes.

Components
The NAEP HSTS collected information on course offerings, grade point averages (GPAs), and course-taking patterns in the nation's schools. Transcript data can be used to show course-taking patterns across years, which may be associated with proficiency in subjects assessed by NAEP.

Transcripts were collected of 12th-grade students who were expected to graduate from high school during the academic year in which the NAEP assessments were administered. Specifically, students included in the 2009, 2005, and 2000 HSTS participated in the NAEP 12th-grade mathematics or science assessment in 2009, 2005, and 2000,

respectively; students included in the 1998 HSTS participated in the civics, reading, and writing assessments in 1998; students included in the 1994 HSTS participated in the geography, reading, and U.S. history assessments in 1994; students included in the 1990 HSTS participated in the mathematics, science, and reading assessments in 1990; and students included in the 1987 HSTS participated in the 1986 long-term trend NAEP assessments in mathematics and science.

Periodicity
High school transcript studies have been conducted by NCES in conjunction with NAEP since 1987. NAEP collected transcript data in 1987, 1990, 1994, 1998, 2000, 2005, and 2009.

2. USES OF DATA

Transcript data are frequently used to address national policy concerns. Transcripts are also an important component of administrative records at the local level. National transcript data provide some benchmarks against which states and localities can assess their students' academic achievement [or performance].

3. KEY CONCEPTS

Some key terms related to the NAEP high school transcript studies are defined below.

Advanced Placement (AP). The AP Program is designed to prepare students to take the Advanced Placement examinations given by the Educational Testing Service (ETS). Students who pass these tests may be given credit and/or be exempted from requirements in colleges and universities based on their scores. Colleges and universities make their own rules regarding what tests to accept and the scores needed for credit or exemptions.

Carnegie Unit. A factor used to standardize credits indicated on transcripts across the studies. The Carnegie unit is a strictly time-based reference for measuring secondary school attainment used by American universities and colleges. A single Carnegie unit is equal to 120 hours of classroom time over the course of a year at the secondary American high school level. Strictly speaking, this breaks down into a single 1-hour meeting on each of five days per week for a total of 24 weeks per year. However, knowing that classes usually meet for 50 minutes yields a value of 30 weeks per year. A semester (one half of a full year) earns 1/2 Carnegie unit.

Catalog. A document compiled by a school or a district listing all available courses that are offered by the school and a description of those courses. Curriculum specialists review catalogs and use them to determine the appropriate Classification of Secondary School Courses (CSSC) code for each course.

Classification of Secondary School Courses (CSSC). A coding system employed for the purpose of standardizing high school transcripts. The CSSC is a modification of the Classification of Instructional Programs (CIP) code used for classifying college courses and contains approximately 2,300 course codes. Each CSSC code contains six digits. The first two digits identify the main program area, the second two digits represent a subcategory of courses within the main program area, and the final two digits define the specific course. For example, for the CSSC code 400522, the first two digits (40) define the Physical Sciences program area, the middle two digits (05) define the Chemistry subcategory, and the final two digits (22) define the course Advanced Chemistry.

Course offerings file. A high school transcript study data file that provides a comprehensive list of the courses offered in the schools included in the study. A CSSC code is associated with each course title.

Grade Point Average (GPA). The GPA is the average of the points earned for all courses taken and does not include courses where the graduate did not receive a grade (audited or pass/fail courses). Each letter grade is assigned a set number of points (A = 4, B = 3, C = 2, D = 1, F = 0), which are adjusted based on the number of Carnegie units earned in each course. GPAs are not standardized for differences in grading practices across schools and teachers, nor are they adjusted for International Baccalaureate (IB), AP, or honors courses.

International Baccalaureate (IB). A nonprofit educational foundation program consisting of a comprehensive 2-year international curriculum that allows students to fulfill the requirements of their national or state education systems.

Secondary School Taxonomy (SST). The framework initially used by the high school transcript studies for analyzing transcript data. The taxonomy divides high school coursework into three distinct curricula: academic, vocational, and personal/other.

Taxonomy. The classification of items into larger categories. In the high school transcript studies, the items are specific secondary school courses (e.g., composition, first-year algebra, AP biology, American government) that are classified into course subject categories, as organized according to the SST, which is based on course content and level.

Tests and Honors file. A data file providing a list of honors and standardized test results, including SAT and ACT scores, that are found in the transcripts.

Transcript. A student's secondary school record containing courses taken, grades, graduation status, and attendance. In addition, it often includes scores from assessments, such as the PSAT, SAT, ACT, and a list of honors.

Transcript file. A data file providing a complete list of all courses appearing in the transcripts of students sampled in the study.

4. SURVEY DESIGN

Target Population
The target population for the NAEP HSTS includes all 12th-graders enrolled in public and private schools in the United States.

Sample Design
The NAEP high school transcript studies were conducted using similar two-stage sample designs that consisted of the selection of schools and then the selection of students within schools. The sample designs for different collections of the NAEP HSTS are discussed below, beginning with the most recent collection.

The 2009 High School Transcript Study. The sample for the 2009 HSTS was designed to achieve a nationally representative sample of public and private high school graduates from the Class of 2009. The sample for public schools was a subset of the 12th-grade public school sample for the 2009 NAEP mathematics and science assessments in order to eliminate the oversampling of public schools in the NAEP study. Private schools were not oversampled in the 2009 NAEP assessments; therefore, all private schools sampled for NAEP were also sampled for HSTS. The 2009 HSTS collected approximately 37,600 transcripts (from a sample of 41,200) from about 130 private schools and 610 public schools.

For those HSTS sample schools that participated in the NAEP assessment, all grade 12 students who were assigned operational mathematics and science booklets were included in the HSTS student sample regardless of whether the student completed the operational mathematics and science booklets (some students did not participate due to absence or exclusion). More than 80 percent of the students in the 2009 HSTS also participated in NAEP. Some students selected for the HSTS sample did not participate (for example, students whose school did not provide a complete, eligible transcript or who did not graduate in 2009).

For those HSTS sample schools that did not participate in the NAEP assessment, no students were included in the HSTS student sample. This was a departure from previous studies where a random sample of 50 students was selected from schools that were selected for NAEP but did not participate. This new procedure was adopted in 2009 to simplify data collection procedures for HSTS.

The 2005 High School Transcript Study. The sample for the 2005 HSTS was designed to achieve a nationally representative sample of public and private high school graduates in the class of 2005. For public schools, the HSTS sample was the 12th-grade public school sample for the 2005 NAEP mathematics and science assessments; that is, the HSTS sample included every eligible school contacted to participate in the 2005 NAEP 12th-grade public school assessments, regardless of whether they actually participated in NAEP. For private schools, the HSTS sample was a subsample of the 2005 NAEP 12th-grade private school sample for the mathematics and science assessments, because private schools were oversampled in the 2005 NAEP. Over 26,000 transcripts from graduates were collected for the 2005 HSTS from a sample of about 640 public schools and 80 private schools.

Only schools eligible for participating in the main national NAEP mathematics and science assessments were eligible for the HSTS. Schools involved with the state-level NAEP assessment or the NAEP trend assessments were not eligible. Within eligible schools, the HSTS used the same samples of students who received the NAEP mathematics and science assessments. For schools that were selected for NAEP but did not participate, graduates were randomly selected. Approximately 94 percent of HSTS sampled students were enrolled in schools that also participated in the NAEP assessments, and about 63 percent of the participating HSTS students also participated in NAEP.

The 2000 High School Transcript Study. The 2000 HSTS school sample comprised all 320 12th-grade public schools and a subsample of the 620 12th-grade private schools selected for the 2000 NAEP. The objective of private school subsampling was to reverse the oversampling of private schools in the 2000 NAEP so that private school students in the 2000 HSTS would be represented in proportion to their prevalence in the general 12th-grade student population.

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, 21,090 actually graduated by October 2000 and 2,360 did not. From the 21,090 graduates, 20,930 transcripts were collected and processed. That is, 99 percent of the transcripts of eligible students were obtained.

The 1998 High School Transcript Study. The 1998 HSTS sample was nationally representative at both the school and student levels. The sample was composed of

schools selected for the NAEP main sample that had 12th-grade classes and were within the 58 primary sampling units (PSUs) selected for the HSTS. A subsample of 320 schools, consisting of 270 public schools and 50 private schools, was selected from the eligible NAEP sample. Some of these schools were NAEP replacement schools, which were asked to participate in the transcript study in order to maintain as many links as possible with NAEP scores. Of the 320 schools in the original HSTS sample, 260 participated in the study, and 230 in both NAEP and HSTS thus maintaining links between students' transcript and NAEP data.

A total of 28,760 students were selected for inclusion in the 1998 HSTS. Of these, 27,180 students were from schools that maintained their NAEP administration schedules and were identified by their NAEP booklet numbers. Another 500 students were from schools that participated in NAEP but had lost the link between student names and NAEP booklet numbers, and 1,080 were from schools that did not participate in NAEP. Of the *28,760* students in the original sample, 25,250 were deemed eligible for the transcript study, and 24,220 transcripts were collected and processed.

The 1994 High School Transcript Study. The 1994 HSTS school sample was nationally representative of all high schools in the United States. The sample was composed of a subsample of 330 public schools and 50 private schools drawn from the lists of eligible NAEP public and private schools. One of these schools had no 12th-grade students and was not included in the HSTS. Of the 379 remaining schools, 340 participated in the 1994 HSTS.

The student sample was representative of graduating seniors from each school. Thus, it included only those students whose transcripts indicated that they had graduated between January 1, 1994, and November 21, 1994. Approximately 90 percent of students in the 1994 HSTS also participated in the 1994 NAEP; the remaining students were sampled specifically for the transcript study, either because their schools did not agree to participate in the 1994 NAEP or because they participated in NAEP but did not retain the lists linking NAEP IDs to student names. The 1994 HSTS also included special education students who were excluded from the 1994 NAEP. High school transcripts were collected for 25,500 students from an eligible sample of 26,050 students.

The 1990 High School Transcript Study. The sample of schools was nationally representative of schools with a grade 12 or having 17-year-old students. The sample of students was representative of graduating seniors from each school. These students attended 330 schools that had previously been sampled for the 1990 NAEP. Approximately three-fourths of the sampled students had participated in the 1990 NAEP assessments; the remaining students attended schools that did not participate in NAEP or did not retain the lists linking student names to NAEP IDs. The 1990 HSTS also included special education students who had been excluded from the 1990 NAEP. In spring 1991, transcripts were requested for 23,270 students who graduated from high school in 1990; *21,610* transcripts were received.

The 1987 High School Transcript Study. The 1987 HSTS was conducted in conjunction with the 1986 long-term trend NAEP assessments. Thus, the HSTS school sample was a nationally representative sample of 500 secondary schools that had been selected for the 1986 long-term trend NAEP assessments. The HSTS student sample represented an augmented sample of participants in the 1986 NAEP who were enrolled in the 11th grade and/or were 17 years old in the 1985–86 school year and who successfully completed their graduation requirements prior to fall 1987. The 1987 HSTS included (1) students who were selected and retained for the 1986 NAEP assessment; (2) students who were sampled for the 1986 NAEP, but were deliberately excluded due to severe mental, physical, or linguistic barriers; and (3) all students with disabilities attending schools selected for the 1986 assessment. Of the 500 schools selected for the HSTS, 430 participated. There were 35,180 graduates in the sample, for whom 34,140 transcripts were received.

Data Collection and Processing

The data collection portion of HSTS 2009 took place in two phases. During Phase 1, from September 2008 through March 2009, field workers contacted schools in their region by phone and in person. Phase 1 involved introducing the study to the sampled schools, obtaining school and course information necessary to understand the content of the transcripts that would be collected during Phase 2, flagging sampled students' records to facilitate Phase 2 data collection, and arranging a return visit to the school to collect transcripts from graduating students. Phase 2 took place from the end of the school year (May) through October 2009. During this phase, selected field workers visited the schools to collect the sampled students' current transcripts.

Sample transcripts. Because transcript format varies greatly among school districts throughout the country, NAEP field workers obtained from each school a transcript representative of either a 2009 graduate's transcript, if available, or of a previous graduate during the initial visit to the school. The field workers marked the transcript to indicate where the key transcript information was found on the transcript and how information regarding course level was coded. Attached to the marked-up transcript was a Transcript Format Checklist indicating the key transcript

information and whether that information was found and, if so, whether it was marked on the school's transcripts.

HSTS School Information Form (SIF). The SIF was completed by the field worker with input from the school coordinator and contained general school information as well as the following information: sources of information within the school (if needed to complete HSTS 2009 data collection); graduation requirements; grading practices at the school; format of the school's transcripts; and name and position of the school's HSTS 2009 coordinator who helped complete the form.

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. The SIF was then forwarded to the home office for data processing along with the other preliminary materials as described above.

School Background Questionnaire. The School Background Questionnaire is a NAEP 2009 questionnaire that collects 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 2009.

SD and ELL Background Questionnaires. The questionnaires that NAEP 2009 used to collect information from school staff about students with disabilities (SD) and English language learners (ELL) are called the SD Background Questionnaire and ELL Background Questionnaire, respectively. Schools were asked to have the person most knowledgeable about an SD or ELL student complete the questionnaire(s). 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.

Student Background Questionnaire. The Student Background Questionnaire used in NAEP 2009 collects information on students' demographic characteristics, classroom experiences, and educational support. The 10-minute questionnaires are completed voluntarily by students and are located at the end of the assessment booklets.

Transcript Request Form (TRF). Once graduation information was posted on transcripts, a field worker returned to the school to obtain the requested transcripts. At that time, the field worker used a Transcript Request Form (TRF) to obtain basic information about the sampled students that was not available from NAEP 2009 data files. In addition to student name and NAEP ID, it contained columns for entering graduation status, gender, birth month and year, race/ethnicity, SD status, ELL status, Title I participation, and National School Lunch Program participation for each listed student.

After completing the TRF 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. 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.

Transcripts. When the request for transcripts was filled, the field worker reviewed the transcripts to ensure that a transcript had been received for each 12th-grade student selected for the operational mathematics or science portion of the NAEP 2009 assessment, whether or not that student had graduated. Even though nongraduate transcripts were not included in HSTS, each student graduation status needed to be accounted for and verified so that weighting could be done correctly. Each transcript was checked for eligibility, understandability (e.g., all the codes on it were 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 student's NAEP ID. If a school did not have a student's transcript on file, the field worker completed a Documentation of Missing Transcripts form to explain the reasons the school gave for any missing transcripts.

Data processing. The main NAEP study provided HSTS staff with data files for schools and students included in NAEP 2009. The questionnaires used to gather information about schools and students for HSTS 2009 were the School Background Questionnaire, Students with Disabilities (SD) Background Questionnaire, English Language Learners (ELL) Background Questionnaire, and Student Background Questionnaire. Unlike previous HSTS studies, the 2009 NAEP survey data fully overlapped with HSTS; therefore, no additional NAEP survey data were processed for HSTS.

To process non-questionnaire data collected as part of HSTS 2009, project staff used WesTes, a custom-built Structured Query Language (SQL) server application specifically designed for processing large-scale transcript-based studies in an accurate and efficient manner. WesTes stores most of the school and graduate information collected for the study in a single integrated relational database. It is used to ensure that the data collected by HSTS 2009 is properly tracked and to assist the data entry and coding personnel in the prompt and accurate completion of their tasks. This stage includes entering data from the SIFs and using WesTes for receipt control, catalog coding, transcript entry, and title matching.

Estimation Methods

Two types of weights were created in the 2009 HSTS, unlinked weights and linked weights. Unlinked weights are student weights designed to compute estimates of all high school graduates in the HSTS sample. Linked weights are student-level weights designed to compute estimates of high school graduates in the HSTS sample linked to a particular National Assessment of Educational Progress (NAEP) assessment. Imputation was done for missing data in the 1994, 1998, and 2000 HSTS conducted in conjunction with NAEP.

Weighting. There is one set of unlinked weights, and two sets of linked weights, one for each assessment subject (mathematics and science). Each set of weights includes a survey weight, used to produce point estimates, and replicate weights, used to compute variances for point estimates. All sets of these weights are designed to estimate variables for all graduates. The type of weight that should be used depends upon the type of data the user is analyzing. For example, in estimating the GPA of graduates, the set of unlinked weights would be used. If, however, the user wishes to explore the relationship between NAEP mathematics scores and GPA, the user would use the set of linked weights for mathematics, because these estimates must be based on the subsample of all HSTS graduates who also took the NAEP mathematics assessment. Similarly, if the user wishes to explore the relationship between NAEP science scores and GPA, the user would use the set of linked weights for science, because these estimates must be based on the subsample of all HSTS graduates who also took the NAEP science assessment.

Unlinked weights. The HSTS unlinked weights reflect the probability-sampling scheme used to arrive at the sample of students for whom transcripts were requested. The HSTS unlinked weights were constructed without regard to the NAEP participation or nonparticipation status of schools and students. The HSTS unlinked student-level weight contains five components: the student base weight, a school-level weight-trimming adjustment, a school-level nonresponse adjustment, a student-level nonresponse adjustment, and a student-level weight-trimming adjustment.

Linked weights. Two sets of NAEP linked weights were computed, one for each assessment (mathematics and science). The linked weights were computed using a weighting procedure similar to the HSTS unlinked weights. Each assessment sample represents the full population, so each of the two sets of NAEP linked weights aggregates separately to the population totals. The HSTS linked student-level weight also contains five components: the student base weight, a school-level weight-trimming adjustment, a school-level nonresponse adjustment, a student-level weight-trimming adjustment, and a student-level nonresponse adjustment.

Imputation. In the 1994, 1998, and 2000 HSTS, it was not possible to obtain transcripts for a small percentage of high school graduates. In addition, some transcripts were considered unusable, since the number of standardized credits shown on the transcript was less than the number of credits required to graduate by the school. Thus, an adjustment was necessary in the weights of high school graduates with transcripts to account for the missing and unusable transcripts. To do this adjustment correctly, it was necessary to have the complete set of high school graduates, with or without transcripts. Students who did not graduate were not included in this adjustment, but they were retained in the process for poststratification. There were a few students, however, for whom no transcripts were received and whose graduation status was unknown. Among these students a certain percentage were imputed as graduating, based on the overall percentages of high school graduates. The remaining students were imputed as nongraduating. The imputation process was a standard (random within class) hot-deck imputation.

For each student with unknown graduation status, a "donor" was randomly selected (without replacement) from the set of all students with known graduation status from the same region, school type, race/ethnicity, age class, school, and sex, in hierarchical order. The two race/ethnicity categories were (1) White, Asian, or Pacific Islander; and (2) Black, Hispanic, American Indian, or other. There were two age classes (born before October 1979; born during or after October 1979). Each student with known graduation status in a cell could be used up to three times as a donor for a student in the same cell with unknown graduation status. If insufficient donors were available within the cell, donors were randomly selected from students in another cell with similar characteristics to the cell in question. At the least, a donor had to be from the same region, type of school, race category, and age category.

Future Plans

As the NAEP program moves toward more computer-based data collection methods, it will also incorporate digital transcript collection methods in the NAEP HSTS. The next HSTS collection is scheduled for 2019, and will be a hybrid collection of traditional paper-based, and digital transcripts. The sample design has not been finalized, and is expected to be similar to 2009; however it could include state or district level designs as well as the national level design. HSTS will continue its linkage with the NAEP mathematics and science assessments, and courses will be coded using the School Courses for Exchange of Data (SCED), a common course coding

system already in use among other NCES transcript studies.

A feasibility study is being conducted to assess the efficacy of collecting digital transcripts of middle grade students from selected jurisdictions whose schools were selected for the NAEP grade 8 assessments. This study will help prepare the NAEP program for the possibility of future NAEP transcript collections among middle school students.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Similar to other HSTS surveys, 2009 HSTS estimates are derived from a sample, and are subject to sampling error. 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 would be likely to change if it were based on another equivalent sample of individuals drawn in exactly the same manner as the actual sample. Since HSTS 2009 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 underestimate the true variance of survey estimates and should not be used.

HSTS 2009 uses the stratified 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. Each replicate undergoes the same weighting procedure as the full sample so that the jackknife variance estimator reflects the contributions to or reductions in variance resulting from the various weighting adjustments. In all, 62 replicate weights were created on each record in the unlinked and linked HSTS 2009 data sets. This variance replication scheme is the same one traditionally used for NAEP samples.

Nonsampling Error

In any study, estimates are subject to nonsampling errors. For the HSTS, nonsampling errors may include errors due to electronic transcript submission, incorrect completion of the SIF, and human error during catalog and transcript coding. Quality control procedures and processes are conducted during data collection and coding to minimize nonsampling error.

Coverage error. From the schools selected in the HSTS school sample, approximately 3.5 percent of the students were excluded from the NAEP assessment. As the transcript study attempted to collect high school transcripts for all students selected for the assessment, regardless of whether they participated in the NAEP tests, transcripts for as many students were included in the HSTS. A transcript was included in the HSTS analyses if it met the following three requirements: (1) the student graduated with either a standard or honors diploma, (2) the student's transcript contained 16 or more Carnegie credits, and (3) the student's transcript contained more than 0 Carnegie credits in English courses. These additional restrictions reduced the number of 2009 students in the sample from approximately 37,600 graduates to 36,400 graduates.

Unit nonresponse. There is unit nonresponse at both the school and student levels in the HSTS. Unweighted response rates are presented in table HSTS-1 for the transcript studies conducted prior to 2009. Unweighted response rates have not been released for the 2009 cohort. There was a weighted school response rate of 95 percent for the 2009 cohort, and a weighted student response rate of 99 percent.

Table HSTS-1. Unweighted response rates for all eligible schools and students in the NAEP HSTS: Various years, 1987–2009

Year	School response rate	Student coverage rate
1987[1]	87	97
1990	87	93
1994	90	98
1998	88	98
2000	81	99
2005	82	84 [2]
2009	95 [2]	99 [2]

[1] The 1987 HSTS was conducted in conjunction with the long-term trend NAEP assessment, and other years were conducted in conjunction with the main national NAEP study.

[2] Weighted response rate.

SOURCE: Methodology reports for the NAEP High School Transcript Studies. Reports are available at https://nces.ed.gov/pubsearch/getpubcats.asp?sid=031.

An unweighted 82 percent of schools participated in the 2005 high school transcript study, higher than the 81 percent in the 2000 HSTS, but lower than the participation rate in the other NAEP transcript studies. In 2009, as in

other years, the response rates varied with the characteristics of the sample school. For example, a lower weighted response rate was achieved for private schools (81 percent) in 2009.

At the student level, the weighted graduate within-school response rate for the 2009 HSTS was 99 percent. Transcripts were obtained for 84 percent of eligible students in the 2005 HSTS (weighted), which is lower than the student-level response rate in the other transcript studies conducted in conjunction with NAEP. The response rate in the 2000 HSTS, 99 percent (unweighted), was the highest achieved in all six transcript studies conducted before 2009.

Data Comparability

For HSTS 2000, there are two publications containing many comparisons and analyses. The first publication, *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990–2000* (Perkins et al. 2004), is a printed report available from the National Center for Education Statistics (NCES) website and EDPUBS. This report analyzes the changes in course credits earned and GPAs 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 course-taking 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* (Roey et al. 2007), is available on the NCES website. It details the number of credits earned by high school graduates in various school subject fields and by various school and graduate characteristics, including gender, race/ethnicity, academic track, type of locale, school type (public/private), and region of the country. It also contains tables covering graduation requirements, GPAs, and NAEP 2000 mathematics and science assessment scores.

For HSTS 2005, there were a number of publications that expanded the kinds of analyses that had been conducted in the past. The HSTS report *America's High School Graduates: Results from the 2005 NAEP High School Transcript Study* (Shettle et al. 2007) summarized the findings from HSTS 2005. This report presents information about the types of courses 2005 high school graduates took during high school, how many credits they earned, and the grades they received. Information on the relationships between high school records and performance in mathematics and science on the NAEP assessments is also included. Transcripts were collected from a nationally representative sample of 26,000 high school graduates. The 2005 results are compared to the results of earlier transcript studies, and differences among graduates by race/ethnicity, gender, and parent education are examined. Study findings include: 2005 graduates earned approximately three more credits (about 360 additional hours of instruction during their high school careers) than their 1990 counterparts. In 2005, the overall GPA was approximately a third of a letter grade higher than in 1990. Graduates with stronger academic records obtain higher NAEP scores. For example, graduates whose highest mathematics course was geometry or below had average NAEP mathematics scores below the Basic achievement level, while graduates who took calculus had average NAEP scores at the Proficient level. Female graduates' GPAs overall and in mathematics and science were higher than the GPAs of male graduates during each year HSTS was conducted. Among those who took higher-level mathematics and science courses, male graduates had higher NAEP scores than female graduates. Increased percentages of White, Black, Hispanic, and Asian/Pacific Islander graduates completed at least a midlevel curriculum in 2005 compared with 1990. The GPAs of all four racial/ethnic groups also increased during this time. In 2005, both Black and Hispanic graduates were less likely than White graduates to have completed calculus or advanced science courses and to have higher GPAs.

The HSTS 2009 datasets offer new possibilities for data analyses that previous HSTS datasets could not offer. Researchers can analyze relationships between the mean NAEP mathematics and science assessment scores by whether or not graduates took selected mathematics or science courses. Incorporating the HSTS 2009 datasets with the previous HSTS datasets, researchers can track courses by grade level across the transcript studies to determine whether course curricula have changed in the past 2 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.

6. CONTACT INFORMATION

For content information about the HSTS conducted in conjunction with NAEP, contact:

Linda Hamilton
Phone: (202) 245-6360
E-mail: Linda.Hamilton@ed.gov

Mailing Address
National Center for Education Statistics
Institute of Education Sciences
Potomac Center Plaza
550 12th Street, SW
Washington, DC 20202

7. METHODOLOGY AND EVALUATION REPORTS

General

Alt, M.N., and Bradby, D. (1999). *Procedures Guide for Transcript Studies* (NCES 1999-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=199905

Bradby, D., and Hoachlander, G. (1999). *1998 Revision of the Secondary School Taxonomy* (NCES 1999-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=199906

Caldwell, N. (1993). *The 1990 High School Transcript Study Data File User's Manual* (NCES 93-416). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=93416

Legum, S. (1993). The 1990 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1990, 1987, and 1982 High School (NCES 93-423). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=93423

Legum, S., Caldwell, N., Davis, B., Haynes, J., Hill, T.J., Litavecz, S., Rizzo, L., Rust, K., Vo, N., and Gorman, S. (1997). *The 1994 High School Transcript Study Technical Report* (NCES 97-262). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=97262

Nord, C., Hicks, L., Hoover, K., Jones, M., Lin, A., Lyons, M., Perkins, R., Roey, S., Rust, K., and Sickles, D. (2011). *The 2009 High School Transcript Study User's Guide* (NCES 2011–465). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011465

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). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2004455

Roey, S., Caldwell, N., Rust, K., Blumstein, E., Krenzke, T., Legum, S., Kuhn, J., Waksberg, M., Haynes, J., and Brown, J. (2001). *The 1998 High School Transcript Study User's Guide and Technical Report* (NCES 2001-477). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001477

Roey, S., Caldwell, N., Rust, K., Blumstein, E., Krenzke, T., Legum, S., Kuhn, J., Waksberg, M., Haynes, J., and Brown, J. (2001). *The 1998 High School Transcript Tabulations: Comparative Data on Credits Earned and Demographics for 1998, 1994, 1990, 1987, and 1982 High School Graduates* (NCES 2001-498). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001498

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). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005483

Roey, S., Perkins, R., Caldwell, N., Rust, K., and Brown, J. (2007). 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 (NCES 2007-463). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007463

Shettle, C., Cubell, M., Hoover, K., Kastberg, D., Legum, S., Lyons, M., Perkins, R., Rizzo, L., Roey, S., and Sickles, D. (2008). *The 2005 High School Transcript Study: The 2005 High School Transcript Study User's Guide and Technical Report* (NCES 2009-480). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009480rev

Shettle, C., Roey, S., Mordica, J., Perkins, R., Nord, C., Teodorovic, J., Brown, J., Lyons, M., Averett, C., and Kastberg, D. (2007). *America's High School Graduates: Results From the 2005 NAEP High School Transcript Study* (NCES 2007-467). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007467

Thorne, J. (1989). *High School Transcript Study, 1987. Contractor Report (CS 89-447).* National Center for Education Statistics, U.S. Department of Education. Washington, DC. http://files.eric.ed.gov/fulltext/ED315450.pdf

Secondary Longitudinal Studies Program High School Transcript Studies

1. OVERVIEW

Purpose
Since 1982, NCES has conducted four high school transcript studies as part of the Secondary Longitudinal Studies Program. The first NCES-sponsored transcript study was conducted in 1982, as part of the first follow-up to the High School and Beyond Longitudinal Study (HS&B) In 1992, another transcript study was conducted in conjunction with the second follow-up to the National Education Longitudinal Study of 1988 (NELS:88) A third transcript study associated with the longitudinal study series was conducted in 2004-05, as part of the first follow-up to the Education Longitudinal Study of 2002 (ELS:2002). In 2013, a fourth transcript study was conducted as a part of the High School Longitudinal Study of 2009 (HSLS:09).

Components
The 2013 High School Transcript Study. The 2013 High School Transcript Study sought information on course-taking for each student, including course sequence, grades, and credits earned. Data also encompassed information about schools' basic enrollment, testing, and the schools' grading and graduation policies/requirement. Student information collected included type of diploma awarded, cumulative GPA, standardized test scores, participation on specialized programs, and reason for leaving school.

The 2004 High School Transcript Study. The 2004 High School Transcript Study (conducted as part of the first follow-up of ELS:2002) sought information about course-taking from the student's official high school record, including courses taken while attending secondary school, information on credits earned, year and term a specific course was taken, and final grades. When available, other information was collected, including dates enrolled, reason for leaving school, and standardized test scores. Once collected, the information was transcribed and linked back to the student's questionnaire or assessment data collected in ELS:2002. Due to the size and complexity of the file, and because of reporting variations by school, additional variables were constructed from the raw transcript file. These composite variables include standardized grade point average (GPA), high school academic program, total credits earned by subject, and others.

The 1992 High School Transcript Study. The 1992 High School Transcript Study (conducted as part of the second follow-up of NELS:88) collected detailed information about the types of degree programs, periods of enrollment, majors or fields of study, specific courses taken, grades and credits attained, and credentials earned.

The 1982 High School Transcript Study. The 1982 High School Transcript Study (conducted as part of the first follow-up of HS&B) collected information on the course-taking behavior of members of the 1980 sophomore cohort throughout their four years of high school. Data include a six-digit course number for each course taken; course credit, expressed in Carnegie units (a standard of measurement that represents one credit for the completion of a 1-year course); course grade; year course was taken; GPA; days absent; and standardized test scores.

Periodicity
High school transcript studies have been conducted by NCES as part of the Secondary Longitudinal Studies Program since 1982. Transcript studies associated with the Secondary Longitudinal Studies Program were conducted in 1982, 1992, 2004, and 2013.

2. USES OF DATA

Transcripts include information that is considered to be the official and fixed record regarding student course taking behavior. It represents a record of courses taken by the student. This information can be used to examine course-taking patterns of students and to predict future education outcomes.

3. KEY CONCEPTS

Some key terms related to the Secondary Longitudinal Studies Program high school transcript studies are defined below.

Advanced Placement (AP). The AP Program is designed to prepare students to take the Advanced Placement examinations given by the Educational Testing Service (ETS). Students who pass these tests may be given credit and/or be exempted from requirements in colleges and universities based on their scores. Colleges and universities make their own rules regarding what tests to accept and the scores needed for credit or exemptions.

Carnegie Unit. A factor used to standardize credits indicated on transcripts across the studies. The Carnegie unit is a strictly time-based reference for measuring secondary school attainment used by American universities and colleges. A single Carnegie unit is equal to 120 hours of classroom time over the course of a year at

the secondary American high school level. Strictly speaking, this breaks down into a single 1-hour meeting on each of five days per week for a total of 24 weeks per year. However, knowing that classes usually meet for 50 minutes yields a value of 30 weeks per year. A semester (one half of a full year) earns 1/2 Carnegie unit.

Catalog. A document compiled by a school or a district listing all available courses that are offered by the school and a description of those courses. Curriculum specialists review catalogs and use them to determine the appropriate Classification of Secondary School Courses (CSSC) code for each course.

Classification of Secondary School Courses (CSSC). A coding system previously employed by the high school transcript studies for the purpose of standardizing high school transcripts. The CSSC is a modification of the Classification of Instructional Programs (CIP) code used for classifying college courses and contains approximately 2,300 course codes. Each CSSC code contains six digits. The first two digits identify the main program area, the second two digits represent a subcategory of courses within the main program area, and the final two digits define the specific course. For example, for the CSSC code 400522, the first two digits (40) define the Physical Sciences program area, the middle two digits (05) define the Chemistry subcategory, and the final two digits (22) define the course Advanced Chemistry.

Course offerings file. A high school transcript study data file that provides a comprehensive list of the courses offered in the schools included in the study. A CSSC code is associated with each course title.

Grade Point Average (GPA). The GPA is the average of the points earned for all courses taken and does not include courses where the graduate did not receive a grade (audited or pass/fail courses). Each letter grade is assigned a set number of points (A = 4, B = 3, C = 2, D = 1, F = 0), which are adjusted based on the number of Carnegie units earned in each course. GPAs are not standardized for differences in grading practices across schools and teachers, nor are they adjusted for International Baccalaureate (IB), AP, or honors courses.

International Baccalaureate (IB). A nonprofit educational foundation program consisting of a comprehensive 2-year international curriculum that allows students to fulfill the requirements of their national or state education systems.

School Courses for the Exchange of Data (SCED). The current classification system used by the high school transcript studies for prior-to-secondary and secondary school courses. It can be used to compare course information, maintain longitudinal data about student coursework, and efficiently exchange course-taking records. SCED is based on a five-digit Course Code that provides a basic structure for classifying course content. Additional SCED elements and attributes provide descriptive information about each course.

SCED is updated and maintained by a working group of state and local education agency representatives who receive suggestions and assistance from a wide network of subject matter experts at the national, state, and local levels. As a result, SCED is designed to be flexible enough that education agencies can modify it to meet their needs.

Secondary School Taxonomy (SST). The framework initially used by the high school transcript studies for analyzing transcript data. The taxonomy divides high school coursework into three distinct curricula: academic, vocational, and personal/other.

Taxonomy. The classification of items into larger categories. In the high school transcript studies, the items are specific secondary school courses (e.g., composition, first-year algebra, AP biology, American government) that are classified into course subject categories, as organized according to the SST, which is based on course content and level.

Tests and Honors file. A data file providing a list of honors and standardized test results, including SAT and ACT scores, that are found in the transcripts.

Transcript. A student's secondary school record containing courses taken, grades, graduation status, and attendance. In addition, it often includes scores from assessments, such as the PSAT, SAT, ACT, and a list of honors.

Transcript file. A data file providing a complete list of all courses appearing in the transcripts of students sampled in the study.

4. SURVEY DESIGN

Target Population

The target population for the high school transcript studies conducted as part of the Secondary Longitudinal Studies Program included all students in public and private schools who participated in previous data collections. For example, the target population for the 2004 high school transcript study included students who had been in-school sophomores in the 2001–02 school year, participated in both the base-year and first follow-up interviews, completed the mathematics assessment in the base-year and first follow-up interviews, and had complete transcript information for the 2002–03 and 2003–04 school years. The 2004 High School Transcript Study included 14,710 of the originally selected sample members of ELS:2002

sophomores in the spring of 2002 who were respondents in both the base-year and first follow-up interviews.

Sample Design
Sample design is essentially similar across the various administrations of the high school transcript studies: multistage, stratified, and clustered.

The 2013 High School Transcript Study. This study was conducted as part of the HSLS:09. A total of 3,028 out of 4,249 schools (base-year and transfer) submitted transcripts, resulting in an unweighted school participation rate of 71 percent. A total of 910 out of 944 base-year schools submitted transcripts, resulting in a 96 percent participation rate. Among the transfer schools, 2,118 of 3,305 schools submitted transcripts, representing an unweighted transfer school participation rate of 64 percent.

School Information Pages (SIPs) and course catalogs were collected from both base-year and transfer schools. As previously stated, the SIP asked for information about the school's grading and graduation policies/requirements. Seventy-six percent or 3,248 schools provided SIP information. Course catalogs were requested from base-year schools that had not provided them in the first follow-up and from all transfer schools. At least one course catalog was received or was able to be retrieved from school/district websites for 3,154 schools (74 percent).

Transcripts were collected for all student sample members who participated in either the base-year data collection, first follow-up data collection, or both. In addition, transcripts were collected for a subset of students who were incapable of participating in prior rounds due to physical or cognitive disabilities Of the 25,167 transcripts requested, 21,928 were received (88 percent). The student coverage rate was 94 percent, where 21,928 of the 23,415 students have complete or partial data. Student coverage is defined as having any transcript data provided by a school. Characteristics of student coverage are based on the base-year school because students may have attended schools with varying demographics throughout their high school career.

The 2004 High School Transcript Study. This study was conducted as part of the ELS:2002 first follow-up in 2004. A total of 1,550 out of 1,950 schools participated in the request for transcripts for an unweighted participation rate of 79 percent. The base-year school weighted response rate was 95 percent. The course offerings response rate for base-year schools was 88 percent, with at least some transcript information available for 91 percent (weighted) of the entire student sample (14,920 out of 16,370 students).

Transcripts were collected for sampled students from the school that they were attending in the base year (which was the only school for most sample members) and from their last school of attendance (if it was learned during the first follow-up student data collection that they had transferred). Incomplete records were obtained for sample members who had dropped out of school, had fallen behind the modal progression sequence, or were enrolled in a special education program requiring or allowing more than 12 years of schooling. For freshened students, transcripts were only collected from their senior-year school. Transcripts were collected for regular graduates, dropouts, early graduates, and students who were homeschooled after their sophomore year.

The 1992 High School Transcript Study. This transcript study was conducted as part of the NELS:88 second follow-up. Including the *freshened* students, there were 4,788 locations. Once non-school locations associated with dropouts, early graduates, institutionalized sample members, home study students, and unlocatables were subtracted from the total, there were 2,260 school sites. Teacher, school administrator, and student transcript components were limited to a maximum of 1,500 schools. For this reason it was still necessary to select a sample of schools, although the students falling outside that sample would not be excluded from the study. For students in the 1,500 schools selected, the full range of data—student, parent, teacher, school administrator, and transcript data—were collected, and these were called contextual schools*;* for the students in a noncontextual school (i.e., in a school not selected to be part of the 1,500), only student and parent data were collected. Transcript data were also collected for all dropouts, early graduates, and twelfth-grade sample members ineligible for the base year, first follow-up, and second follow-up surveys owing to a language, physical, or mental barrier. All schools identified as having four or more first follow-up sample members enrolled were included in the school-level sample with certainty (probability = 1.0), and random samples were selected for retention from schools identified as having three first follow-up members (probability = 0.75), two first follow-up members (probability = 0.65), and one first follow-up member (probability = 0.32). In the end, 1,370 contextual schools were contacted to participate in the transcript study. An additional 470 noncontextual schools were contacted to provide information on students' transcripts, but the full instrument of school questions was not asked.

Of the 1,840 schools selected for the study (including both contextual and noncontextual schools), 1,540 participated. Transcripts were requested for 19,320 students, and 17,290 transcripts were received.

The 1982 High School Transcript Study. The first transcript study was a component of the HS&B first follow-up. The study included students from 1,900

secondary schools—1,000 HS&B sampled schools and 900 schools to which students selected for the transcript survey had transferred (and for which no data collection activities other than transcript collection were carried out). Of these 1,900 schools, 1,720 provided transcripts.

The total student sample size was 18,430 students. Of the 1980 sophomores selected for the HS&B first follow-up, 12,310 cases were retained in the study sample with certainty—12,030 cases in the probability sample plus 280 nonsampled co-twins. In addition, 6,120 cases were systematically subsampled from the 17,700 remaining first follow-up selections, with a uniform probability of approximately 0.35. Transcripts were collected for 15,940 of the 18,430 students.

Data Collection and Processing

The data collection and processing procedures are similar across the four transcript studies conducted as part of the Secondary Longitudinal Studies Program. The data collection procedures of the 2004 and 2013 High School Transcript Study are discussed to illustrate the data collection process.

Data Collection. The HSLS:09 transcript data collection commenced in September 2013 with the mailing of materials to all base-year schools. Transfer schools (e.g., a school attended by a sample member who transferred out of their base-year school) identified during the HSLS:09 first follow-up and the 2013 Update received the initial request for transcripts in October 2013. Control systems were designed to manage the transcript and other data requested from schools. Institution contactors (ICs) served as liaisons to schools that provided the requested materials through a variety of possible submission methods of possible submission methods, including a study website. As new transfer schools were identified during the transcript data collections, transcript request materials were sent to these schools if a complete transcript record had not already been obtained from another school. The materials were designed to request transcript data and guide school personnel in the preparation of transcripts and related documents. The materials also directed school staff to the study website where additional information about the transcript collection could be obtained. Student-level information was provided on the transcripts, while school-level information was collected separately on the School Information Page (SIP). The instructions for preparing student transcripts and SIP data requested that transcripts be prepared for the students listed on the secure study website. The transcripts could be uploaded via the secure study website, faxed to a secure fax number, sent as an encrypted attachment by e-mail, or sent by FedEx (redacted transcripts). The information about the school's grading and graduation policies/requirements collected on the SIP could be completed online or by hardcopy. If the SIP was missing any key information such as the grading scale or term system, ICs followed up with schools to obtain this information or searched for it online. SIP information was used to provide context for transcript data collected from each school.

In addition to collecting data from base-year schools, transcript data were collected from the transfer schools of students who left their base-year high school. Transfer students, and the schools to which they transferred, were identified at several points in the HSLS:09 data collection process. These time periods included the enrollment status update completed by the schools in the fall of 2011, questionnaire data provided by students and parents in the spring of 2012, and other contact updates with students and parents, most recently being the spring 2013 update. Schools were also identified through the analysis of transcripts that were submitted. A complete transcript was defined as having at least five courses in each of the four grades plus a graduation date. Users need to be cautious when including those sample members with incomplete high school transcript information. This situation occurs when the data are either missing or censored. Missing transcript information may result from unit nonresponse from the school, inability to obtain multiple transcripts for certain students who have transferred, or school recordkeeping errors or inconsistencies.

The ELS:2002 transcripts were collected from sample members in late 2004 and early 2005, about six months to one year after most students had graduated from high school. Collecting the transcripts in the 2004–05 school year allowed for more complete high school records. Transcripts were collected from the school that the students were originally sampled from in the base year (which was the only school for most sample members) and from their last school of attendance, if it was learned during the first follow-up student data collection that they had transferred. By requesting transcripts and related information for transfer students from a second school, the ELS:2002 transcript study offers the unique advantage of having extensive information on multiple school attendance and, therefore, increased accuracy of enrollment histories. Incomplete records were obtained for sample members who had dropped out of school, had fallen behind the modal progression sequence, or were enrolled in a special education program requiring or allowing more than 12 years of schooling. For freshened students, transcripts were only collected from their senior-year school. Transcripts were collected for regular graduates, dropouts, early graduates, and students who were homeschooled after their sophomore year.

From December 2004 through June 2005, survey materials were sent to over 2,000 schools. This group included schools that participated either in the base-year or first

follow-up survey and transfer schools that were first contacted regarding ELS:2002 during transcript data collection. Transcripts were not requested from 10 base-year schools, because they had refused to participate in the first follow-up survey. Additionally, transcripts were not requested from one base-year school that had no eligible students. Schools were paid $5 for each transcript. Transcripts were requested for over 16,000 sample members. Included were sample members who were ineligible to participate in the base year or first follow-up because of a physical disability, a mental disability, or a language barrier. Ninety-five schools required explicit consent from sample members or their parents/guardians before releasing transcript information. Of the sample members who attended these schools, about a quarter provided signed release forms.

Two weeks after the survey materials were sent to the school, a follow-up postcard was sent as a reminder to complete the data collection forms and to send the requested materials to the Research Triangle Institute (RTI). If, after an additional week, RTI had not received the materials from the school, assigned institutional contactors (ICs) began telephone prompting to request that the materials be sent as soon as possible. Nonresponding schools contacted during the telephone prompting frequently requested remailing of the data collection materials. During telephone contacts, the ICs also identified any additional requirements the school had for releasing transcripts. Telephone follow-up with schools continued through June 2005. Additional measures were implemented to ensure an adequate response rate. In June 2005, data collection materials were sent to schools that had not yet provided all of the requested transcripts. In addition, in-person visits to nonresponding schools were conducted during April through June 2005 to collect the requested materials or to assist the school transcript preparer in assembling the information. For efficiency, the schools were selected for in-person visits by their proximity to other schools. In-person visits were made only to schools that had not sent transcript materials for any requested sample members.

Data processing. Starting with The High School Longitudinal Study of 2009 (HSLS:09), a new coding system was implemented, departing from the three previous high school transcript studies that used the Classification of Secondary School Courses (CSSC) for course coding. The descriptions of the 2013 high school transcript data processing procedures illustrate the use of the new system, while the 2004 high school transcript data processing procedures illustrate the previous data processing done in the other three transcript studies conducted as part of the Secondary Longitudinal Studies Program.

For the 2013 data processing, the keying and coding system was the web-based data entry application used by the keyer/coders. This system was used for both course catalog and transcript keying and coding. Course catalogs for base-year schools were keyed and coded first, followed by the keying of transcripts for each of those schools. Transcripts from non-base-year schools attended by sample members were coded individually. The School Codes for the Exchange of Data (SCED) was used as the basic classification system; the SCED provides a comprehensive framework for capturing secondary school courses.

The use of the SCED was a departure from previous high school transcript studies, which had used the Classification of Secondary School Courses (CSSC) for course coding. The decision to switch to the SCED was based on several factors, including the additional detail captured with the SCED, the up-to-date nature of the SCED, the widespread adoption of the SCED within the K12 education data community, and advantageous characteristics of the SCED codes for course coding.

More specifically, the SCED provides a five digit course code that categorizes the subject of the course. While this is reminiscent of the CSSC, the SCED also has attributes to capture the sequence of a course (e.g., course one of two), the level (e.g., honors), and the number of Carnegie units available for the course. This more robust coding scheme has been refined through regular meetings of a working group of experts including representatives of state and local education agencies (SEAs and LEAs) who use the SCED. This working group continues to refine the SCED, resulting in a coding scheme that is in line with contemporary course offerings and the needs of data users. The SCED is actively promoted for use by SEAs and LEAs with their own data, and as it is more broadly adopted, the growing familiarity with the SCED can facilitate the use of HSLS:09 data by researchers familiar with the taxonomy.

Finally, the five-digit course code taxonomy of the SCED offered advantages to course coding. The detailed SCED course code descriptions aid in distinguishing between codes and reliably selecting valid codes for courses. The SCED includes 500 fewer total codes than the CSSC, which further aids course coding.

For the 2004 data processing, incoming data collection forms, transcripts, and course catalogs were logged into the survey control system by staff from RTI. Course catalog and transcript data were then entered using a web-based computer-assisted data entry (CADE) system. Course catalogs from ELS:2002 base-year schools were keyed and coded for the preparation of course offerings data. For ELS:2002 base-year schools that provided them,

courses listed in course catalogs were keyed and assigned the appropriate CSSC code before transcript keying and coding. For each catalog course entered, keyer-coders selected an appropriate course code from the CSSC look-up table in the data entry system. All transcripts received from a school were assigned to a single person for keying and coding. Course catalogs from non-base-year schools were not keyed. Data entry of each catalog and transcript was reviewed for accuracy by a supervisor or by a group of keyer-coders trained to perform these reviews.

Procedures for editing, coding, error resolution, and documentation were modeled after the NELS:88 second follow-up transcript component (Ingels et al. 1995). Data entry systems included checks for valid variable ranges and codes, including legitimate missing codes, and CSSC code checks. Sequences of machine edits and visual data inspections were performed. Tasks included supplying missing data, detecting and correcting illegal codes, and investigating and resolving inconsistencies or anomalies in the data. Variable frequencies and cross-tabulations were reviewed to verify the correctness of machine editing.

Estimation Methods

Weighting. The weighting procedures used in the 2013 High School Transcript Study and 2004 High School Transcript Study are presented as an example of the weighting procedures used in transcript studies conducted as part of the Secondary Longitudinal Studies Program.

In the 2009 High School Longitudinal Study, estimates were generated for the target populations with a set of analytic weights and software that accounts for the complex, two-stage sample design. A series of weights have been computed for HSLS:09 to accommodate analyses specific to each round of the study (base year, first follow-up, or 2013 Update and high school transcript collection) plus analyses to evaluate change over time.

In the 2004 High School Transcript Study, weight was assigned as follows. First, the first follow-up design weight was used as the starting weight. Next, generalized exponential models (GEMs) were used to compute weight adjustments. Weight adjustments included (1) a nonresponse adjustment to reduce potential bias owing to transcript nonresponse; and (2) a poststratification adjustment to ensure that sums for weights for certain domains had the same totals as those in the first follow-up. The nonresponse adjustment was performed in two stages: (1) at the school refusal stage (e.g., the school refused to provide any transcript); and (2) at the within-school student-level nonresponse stage (see below for more details). Poststratification was performed to keep key estimates consistent with those in the first follow-up. Extreme weights were adjusted, truncated, and smoothed by GEMs as part of the nonresponse and poststratification adjustments rather than as a separate step.

Imputation. Imputation was done for missing data in the high school transcript studies conducted for NELS:88 HS&B, and HSLS:09 as part of the Secondary Longitudinal Studies Program. In the NELS:88 study, imputation was done for missing sex data, using the student's first name to determine sex. In the HS&B study, values were imputed for missing sex and race/ethnicity data. In HSLS:09, imputation was done for missing data from responded records.

Future Plans

High school transcripts were collected in the 2013–14 academic year as part of the HSLS. A second follow-up interview took place in 2016, when most sample members were 3 years beyond high school graduation. The number and timing of future follow-ups beyond 2016 is yet to be determined, although the expectation is that the cohort will be followed at least to age 30, with a questionnaire administration and a postsecondary education transcript collection in 2025–26.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

For the 1982, 1992, 2004, and 2013 high school transcript studies, variance estimation required the Taylor series linearization procedure, which took into account the complex sample design of these surveys, including stratification and clustering. This procedure takes the first-order Taylor series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. For stratified multistage surveys, the Taylor series procedure requires analysis strata and analysis primary sampling units (PSUs). (in ELS:2002, schools are the PSUs). Therefore, analysis strata and analysis PSUs were created in the base year and used again in the first follow-up.

Transcript studies conducted as part of the Secondary Longitudinal Studies Program may also use the balanced repeated replication (BRR) variance estimation procedure or both Taylor series linearization and BRR for variance estimation. For example, in NELS:88 and ELS:2002, variance estimation can be done in two ways: first, with Taylor series linearization using software such as SUDAAN, AM, or STATA when using the Electronic Codebook (ECB) data; or, when using BRR, using the table generator (DAS—Data Analysis System) version of the dataset. Thus, the same estimate can have two different standard errors even within the same study, depending on whether its basis is a Taylor series linearization or BRR.

HS&B used both BRR and the Taylor series and compared the results. These two methods result in very small differences that should not markedly change conclusions about the standard error of an estimate.

Nonsampling Error

Coverage error. Potential sources of undercoverage in the high school transcript studies include (1) incomplete sampling frame data, as no national listing of schools is, or remains for very long, 100 percent complete and accurate; (2) omissions and errors in school rosters; and (3) deliberate exclusion of certain categories of students—such as students with physical or mental disabilities or non-English speakers, who might find it difficult or impossible to complete demanding cognitive tests and questionnaires. The first two sources are thought to have only a very small impact on high school transcript estimates. The most serious potential source is the undercoverage bias due to the exclusion of certain categories of students.

Unit nonresponse. There is unit nonresponse at both the school and student levels in the high school transcript studies conducted as part of the Secondary Longitudinal Studies Program. Response rates for all eligible schools and students are presented in table HSTS-2.

Transcripts were collected from 71 percent (unweighted) of the schools in the 2013 (HSLS:09) transcript study, 79 percent (unweighted) of the schools in the 2004 (ELS:2002) transcript study, 84 percent (unweighted) of the schools in the 1992 (NELS:88) transcript study, and 91 percent (unweighted) of the schools in the 1982 (HS&B) transcript study.

At the student level, transcripts were obtained from 88 percent (unweighted) of eligible students in the 2013 study, 91 percent (unweighted) in the 2004 study, 89 percent in the 1992 study, and 88 percent (unweighted) in the 1982 study.

Table HSTS-2. Unweighted response rates for all eligible schools and students in the Secondary Longitudinal Studies Program High School Transcript Studies: Various years, 1982–2013

Year	School response rate	Student response rate
1982	91	88
1992	84	89
2004	79	91
2013	71	88

SOURCE: Methodology reports for the Secondary Longitudinal Studies Program High School Transcript Studies. Reports are available at https://nces.ed.gov/surveys/slsp/.

Item nonresponse. Rates for item nonresponse have ranged from nonexistent to extremely high, depending on the type of item, across all of the high school transcript studies. As would be expected in transcript studies, course-level items—including school year, term, and grade in which a course was taken; school-assigned course credits; and standardized course grade—have little if any nonresponse. However, nonresponse rates for items such as class size, cumulative GPA, class rank, days absent in each of the four high school years, and standardized test scores (e.g., PSAT, SAT, ACT) are more prevalent.

In the 1992 transcript study, the nonresponse rates for these items ranged from 0 percent for school year to less than 2 percent for the school term in which a course was taken. Incompleteness of actual course data, while considered to be limited, is another source of potential bias in a transcript study. Course data may be incomplete for students who transferred from one school to another. Also, it is difficult to assess the completeness of transcript data for dropouts in the 1982 and 1992 transcript studies because of inconsistencies between enrollment reports of the sample member and the school.

Transcripts often provide other pieces of information that are useful in the analysis of course-taking patterns: days absent in each school year, class rank, class size, month and year student left school, reason student left school (e.g., dropped out, graduated, transferred), cumulative GPA, participation in specialized courses or programs, and various standardized test scores (e.g., PSAT, SAT, ACT). While nonresponse rates for participation in specialized courses or programs (2 percent) and month/ year/reason student left school (less than 4 percent) were quite low in the 1992 transcript study, nonresponse rates for the other items were very high: 18 percent for class size; 22 percent for cumulative GPA; 23 percent for class rank; 42–44 percent for days absent in each of the 4 high school years; and 67–73 percent for standardized test scores. (Note that although students were asked in a student questionnaire whether and when they planned to take specific tests, some students may not have actually taken the tests; this would, in part, explain the high nonresponse rates for test scores.)

The wide range of item nonresponse rates in the 1992 study is comparable to that in the 1982 study. For example, in the 1982 study, the nonresponse rate was 32 percent for class rank and class size, 41–47 percent for days absent per school year, and 75 percent and above for standardized test scores.

Two key analytic variables are sex and race/ethnicity. Item nonresponse rates for sex have been extremely low: 0 percent in both 1982 and 1992. For race/ethnicity, nonresponse was 0 percent in 1982 and 0.7 percent in 1992. Item nonresponse rates were higher in the 2004

transcript study: 8.8 percent (unweighted) of respondents were missing information on their sex and race/ethnicity.

Measurement error. Possible sources of measurement error in high school transcript studies are differences between schools and teachers in grading practices (e.g., grade inflation), differences in how data are recorded (although efforts are made to standardize grades and course credits for the high school transcript studies), and errors in keying or processing the transcript data (although the system has many built-in quality checks). The amount of measurement error in any survey or study is difficult to determine, and it is unknown for the high school transcript studies. However, because the transcripts are official school records of students' progress, it is reasonable to presume that there is less measurement error than in other types of data collections, particularly those that are self-reported.

Data Comparability

The high school transcript studies conducted by NCES as part of the Secondary Longitudinal Studies Program have both similarities and dissimilarities of design and methodology that raise questions of comparability and may sometimes require analytical adjustments to ensure that comparability is maximized. This section presents four such issues: the comparability of target populations, sample inclusion and exclusion, methodology across studies, and content across studies. For details, please refer to the *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation* (Bozick et al. 2006).

Comparability of target populations. The first comparability issue concerns the comparability of the target population. Comparable analysis samples can be achieved across the high school transcript studies by limiting analysis samples to high school graduates who received regular/standard or honors diplomas and imposing additional restrictions such as earned credit minimums.

HS&B drew a national probability sample of high schools, as well as the sophomores and seniors within those schools, as of the 1980 spring term. By 1982, the school sample was no longer nationally representative (in the strictest sense) because it did not take into account school openings and closings in the two-year period.

Similarly, while the HS&B senior cohort sample in 1980 generalized to the nation's high school seniors, the sophomore cohort in 1982 cannot be said to strictly represent the high school class of 1982. The HS&B sample was never freshened to add 1982 seniors who had no chance of selection two years before. This means that there is a bias in the HS&B 1982 (sophomores 2 years later) sample when it is used to generalize either to 12th-graders or to high school graduates. Seniors who were outside the United States 2 years before or seniors who were not sophomores 2 years before (e.g., seniors who repeated a year or who had a significantly accelerated trajectory) had no chance of selection into the sophomore cohort sample and are not represented within it.

The next two NCES high school cohort longitudinal studies, NELS:88 and ELS:2002, instituted a sample freshening procedure so that they include a nationally representative sample of high school seniors.

Sample inclusion and exclusion. A second issue of comparability across studies concerns student sample inclusion and exclusion, especially with respect to students with disabilities and limited English proficiency.

In HS&B, sample members were classified as ineligible if deemed by their schools unable to complete the HS&B assessment battery owing to disability or lack of proficiency in English. Unfortunately, excluded students and specific reasons for exclusion were not well documented. However, it seems clear that the ineligible students represent the more severely disabled and the least proficient non-English speakers.

While some students were excluded from NELS:88, these exclusions were well documented, and over time the eligibility status of these students was revisited. In ELS:2002, no students were excluded, though for those who could not complete survey forms, only contextual data and transcripts were collected. Also, in ELS:2002, some students received testing accommodations (e.g., extra time to complete the test); these cases are specially flagged.

Limiting 12th-grade high school graduate samples to recipients of regular or honors diplomas and eliminating cases that lack English course credits or that reflect a special education diploma or certificate of attendance largely eliminates the problem of differences in the excluded student population across studies. However, there is the remaining issue of how to identify and study the transcripts of individuals who had mild disabilities and how to compare the results over time. These issues arise because the longitudinal studies sought disability information from multiple respondent populations at multiple points in time. In NELS:88, for example, parents, teachers, students, and school administrators were all used as sources of information related to disability status.

Although some disability information was collected from sophomores' teachers, the primary source of identification for sophomore cohort members with disabilities in ELS:2002 is the Individualized Educational Program (IEP) flag, based on information taken from the sampling records provided by the base-year school, which identifies students in the school with IEPs.

Methodology across studies. In addition to differences in target populations and inclusion criteria, there are differences among NCES high school transcript studies in terms of methodology. First, there is some variation in the statistical procedures used across studies. Overall, this variation is the source of small differences that should not disrupt trend analyses. For example, different methods were used for nonresponse adjustment of weights. In HS&B, weighting cells were constructed based upon the known characteristics of the sample units. ELS:2002 used propensity modeling rather than a weighting cell approach. In NELS:88, a mix of the two approaches is encountered (propensity at the school level, weighting cells at the student level). However, results of nonresponse adjustment tend to be highly correlated, regardless of method. Therefore, these differences should not lead to greatly different estimates.

Content across studies. As curriculum changes, new courses emerge while others fall by the wayside. Therefore, with every transcript study, there is a need to add courses to the CSSC. Additionally, the SST has been revised twice to accommodate changes in the curriculum. From a classification standpoint, adding new subject areas (such as information processing and computer studies) and expanded course offerings (including more AP courses) presents less of a quandary than certain efforts to achieve curriculum integration through interdisciplinary courses. Confining such offerings (e.g., history of mathematics, philosophy of science, psychological anthropology) in one subject category does injustice to certain aspects of the course content, while counting such courses in multiple areas may magnify and distort their impact.

As a result of these changes, many transcript composite variables have also changed over time. For example, with initiatives to seamlessly integrate academic into vocational education, conceptualizations of track or program type have changed. Such differences may reduce ease and simplicity in trend analysis, but are unavoidable features of the need to confront a complex and changing reality. Also, HS&B did not use as refined a system of course classification as did the later studies, which, for example, distinguished courses based on whether they were remedial, regular, or advanced. On the other hand, some new measures developed out of NELS:88, such as the "pipeline" variables, which measure course content level and can be "read into" the other transcript studies.

The major limitation of these changes is that there are few course-taking variables that are directly comparable across studies. For example, only a handful of courses qualified as computer science in the HS&B study. As the number of computer science courses has expanded, any variable based on computer science is not truly comparable across studies because it does not capture the range of courses that have emerged over time. Along with the two revisions of the SST, these changes make it difficult to draw direct comparisons among course-taking variables in the different files. To facilitate some comparisons, ELS:2002 provides six summary measures that have directly comparable variables in NELS:88 and that can be constructed in HS&B by using existing elements. These variables are based on the same CSSC codes.

Analysts interested in comparing course-taking patterns need to examine the CSSC codes available in each study. The CSSC codes are the same across studies, thus facilitating direct comparisons. As noted earlier, the list has evolved and certain subject areas have changed accordingly. Users may want to construct measures in a variety of ways to ensure that their findings are robust with respect to different variable specifications. In addition, analysts should consider changes in subject areas over time when conducting time trend analyses and interpreting findings.

There are many other variables that are typically linkable to transcripts; however, their status for this purpose may sometimes be problematic. For example, in HS&B and NELS:88, race was self-reported and students were asked to mark only one race. In light of the 2000 decennial census and revised race-reporting guidelines issued by the Office of Management and Budget, a new race category was added at the time of ELS:2002. More importantly, ELS:2002 respondents were allowed to mark all applicable races, thus generating a further category—multiracial. Knowing if a respondent who self-identified as Black on the HS&B questionnaire would have self-identified only as Black on the ELS:2002 questionnaire is impossible. To this extent, course-taking trends for Blacks will be more uncertain than if a consistent definition had been maintained.

Test scores are another set of variables typically linked with transcript data that are different across studies. The relationship between course-taking and tested achievement is of interest to researchers, and exploring the relationship between curriculum and assessment results is an interesting area for time series analysis. The NCES transcript studies provide only limited scope for such explorations.

For example, math is the only subject tested regularly across the education longitudinal studies, meaning other subjects matters are less easily examined over time. A further complication with comparative use of assessment data is changes in the measurement scale. Selectively, where content similarities permit, this limitation has been overcome by test linkage, usually item response theory (IRT)-based or equipercentile equating. One could, for example, examine the relationship between course-taking

and gain in the first two years of high school, using the equated 1980, 1990, and 2002 mathematics scores, or one could examine the relationship between course-taking and gain for the periods 1990–92 and 2002–04, since ELS:2002 has been put on the NELS:88 scale. One final option for use of assessment data is to examine change within an effect size metric.

6. CONTACT INFORMATION

For content information on the high school transcript studies conducted as part of the Secondary Longitudinal Studies Program, contact:

Elise Christopher
Phone: (202) 245-7098
E-mail: Elise.Christopher@ed.gov

Mailing Address
National Center for Education Statistics
Institute of Education Sciences
Potomac Center Plaza
550 12th Street, SW
Washington, DC 20202

7. METHODOLOGY AND EVALUATION REPORTS

General

Alt, M.N., and Bradby, D. (1999). *Procedures Guide for Transcript Studies* (NCES 1999-05). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=199905

Bozick, R., Lyttle, T., Siegel, P.H., Ingels, S.J., Rogers, J.E., Lauff, E., and Planty, M. (2006). *Education Longitudinal Study of 2002: First Follow-up Transcript Component Data File Documentation* (NCES 2006-338). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006338

Bradby, D., and Hoachlander, G. (1999). *1998 Revision of the Secondary School Taxonomy* (NCES 1999-06). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=199906

Ingels, S.J., Dowd, K.L., Taylor, J.R., Bartot, V.H., Frankel, M.R., Pulliam, P.A., and Quinn, P. (1995). *National Education Longitudinal Study of 1988 Second Follow-Up: Transcript Component Data File User's Manual* (NCES 95-377). National Center for Education Statistics, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=95377

Ingels, S.J., Pratt, D.J., Herget, D., Bryan, M., Fritch, L.B., Ottem, R., Rogers, J.E., and Wilson, D. (2015). *High School Longitudinal Study of 2009 (HSLS:09) 2013 Update and High School Transcript Data File Documentation* (NCES 2015-036). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2015036

Planty, M., Bozick, R., and Ingels, S.J. (2006). Academic Pathways, Preparation, and Performance: A Descriptive Overview of the Transcripts From the High School Graduating Class of 2003–04 (NCES 2007-316). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007316