

### National Student Clearinghouse

All 2013 postsecondary enrollment and persistence reporting, at the district and campus levels, used data provided by the [National Student Clearinghouse](#) (NSC) on August 15, 2014.

The NSC is a nonprofit organization that has provided postsecondary degree and enrollment verification for more than 17 years. Austin Independent School District (AISD) uses both of these services. More than 3,600 public and private United States institutions, enrolling 98% of all students, participate in NSC.<sup>1</sup> A list of participating colleges and universities, by state, is on the [NSC website](#).

NSC data are updated daily. These data include a unique record for every occurrence of enrollment per semester, quarter, or year. The unit of enrollment time varies by institution and depends on program or degree offerings.<sup>2</sup>

For the most part, records returned as “not found” represent students who did not enroll in a postsecondary institution. However, there are caveats to this matching process. NSC provided a list of reasons a student’s record might not be matched in its database *even though* the student was enrolled in a postsecondary institution:

- The student attended a school that did not report enrollment information to NSC.<sup>3</sup>
- The record was blocked either by the school or at the student’s request. In this case, the student would be counted in the aggregate report provided automatically by NSC, but his or her individual record would not be included in the *Detail Report*.
- The student record was in AISD’s database but not within the search date (i.e., high school graduation date). If a requestor sends a file that includes a non-graduate, the last day that student attended high school may not fall within the search (i.e., the period of time, or range, between the last date of attendance and the day the file is run by NSC). NSC only returns postsecondary records for which enrollment ended *after* the last day of high school attendance.
- A match could not be found with the information provided (e.g., no date of birth, incomplete or inaccurate name).
  - Records missing date of birth will likely be “not matched.”
  - Records missing the last day of attendance cannot be processed.

---

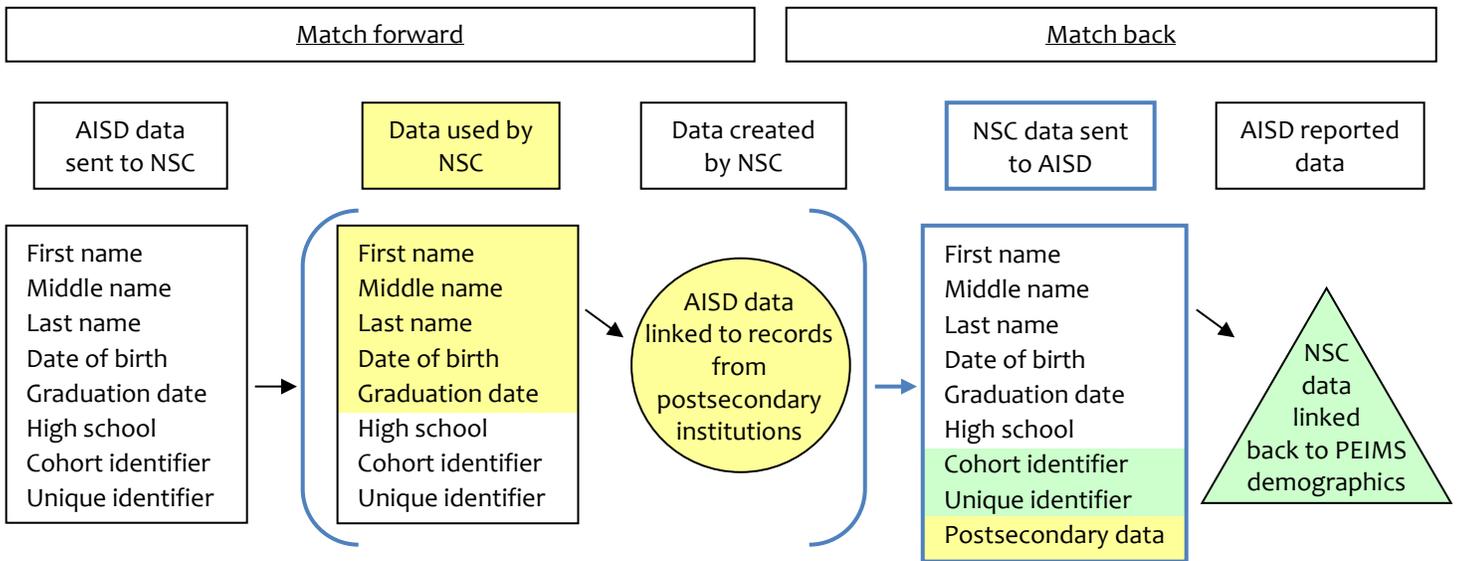
<sup>1</sup> For national and regional coverage rates, see the [NSC Enrollment Coverage Workbook](#).

<sup>2</sup> Colleges can submit limited historical data when they initially join the NSC EnrollmentVerify service. Typically, these data are only for one or two semesters. Institutions that join DegreeVerify can submit historical data at any time.

<sup>3</sup> In 2009, approximately 100 colleges in Texas did not report to NSC. More than half of these institutions were private for-profit schools that offered programs that were at least 2 years but less than 4 years.

- Incomplete or inaccurate names will likely be “not matched.” Examples of an incomplete or inaccurate name include an apostrophe in the name, two last names, and a hyphenated name.<sup>4</sup> Although one of these inaccuracies may not cause a failed match, any combination of these inaccuracies increases the likelihood of a false negative. Colleges do not typically report any kind of punctuation in students’ names.

Figure 1. Relationship Between AISD Student Data and [National Student Clearinghouse](#) (NSC) Data for Postsecondary Analysis<sup>5</sup>



### Data Submission to NSC – Match Forward

**Data sources.** AISD student records were retrieved from current and historical graduate records submitted to the [Texas Education Agency’s \(TEA\)](#) Public Education Information Management System (PEIMS; known internally as PEIMS 203 data tables). Historical records included cohorts 2002 through 2009. The current graduate file included students who graduated at any point during the 2009–2013 academic year. The most recent graduate cohort data table often is not posted at the time of the spring NSC data pull. When graduate data are not available, these are requested from and certified by the PEIMS data manager in Management Information Services.

Data also were retrieved from the student information system (known internally as the ASTU table). This data table included many duplicate student records because of student mobility. Records with the most recent “last day of attendance” were retained for submission to NSC.<sup>6</sup> Data were then linked to the graduate’s file.

<sup>4</sup> We became aware of this limitation when at least one student was not matched because the name we submitted did not match the name his or her postsecondary institution reported. We are working with NSC to develop methods to prevent such matching errors.

<sup>5</sup> The data elements sent to NSC will change for the Class of 2011. NSC has expanded its [StudentTracker](#) product to report outcomes using an array of new student characteristics (e.g., student demographics and assessment outcomes).

<sup>6</sup> All data sources were checked for duplicate records.

Figure 2. Recommended Data Map for Submission to National Student Clearinghouse (NSC)

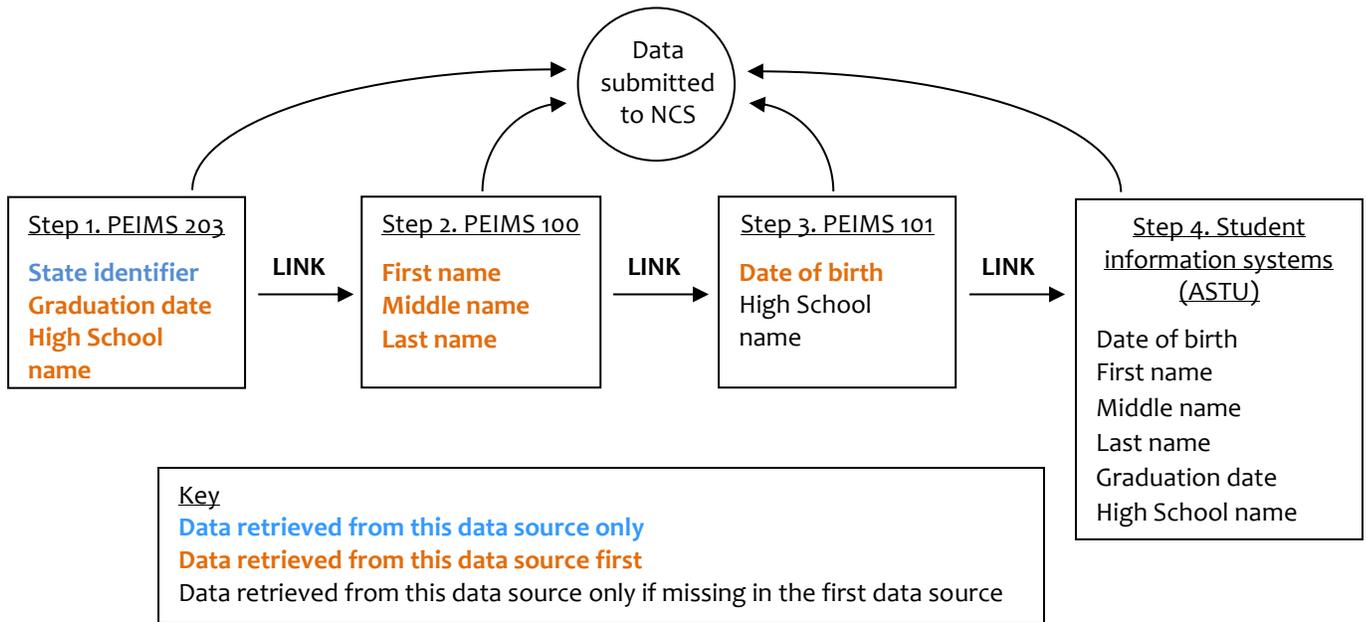


Table 1. Timing of Relevant PEIMS Submissions to Texas Education Agency

Submission	PEIMS record			
	100	101	203	400
October	X	X	X	
January	X	X		X
June	X	X		

**Data modifications.** The campus formerly known as Johnston High School (2002–2003 through 2007–2008) was repurposed as Eastside Memorial High School in 2008–2009 in accordance with state policy for underperforming schools. This campus was then redesigned in 2009–2012 into two campuses: Green Tech and Global Tech High Schools on the Eastside Memorial campus. Eastside seniors for 2012 graduated from Green Tech High School. NSC uses [ACT codes](#) to identify high school campuses. The 2008–2009 transition did not result in a change of the ACT code for this campus location; however, the 2009–2012 transition did.<sup>7</sup> Green Tech and Eastside Memorial High Schools were considered the same for reporting to produce consistent trend data. This method was deemed appropriate because student composition at this campus did not change when the school was repurposed.

All graduate records were submitted to NSC regardless of missing data. Any records that could not be processed were deleted by NSC and were missing in the returned file. These records were

<sup>7</sup> When the school incorporated school-wide programmatic changes and transitioned to a science, technology, engineering, and math school (STEM) in 2009–2012, a new ACT code was assigned to Green Tech (and to Global Tech, although Global did not have any seniors and thus had no graduates).

identified in the match back as “unprocessed.”

*Data sharing.* Data were shared with NSC using the articulated process (i.e., secure file transfer protocol). Data files created in SAS Enterprise Guide (SAS-EG) were exported to Excel for sharing. Data sharing with NSC is not subject to opt-out omissions. NSC does not use these data for research. NSC only uses the data for standard reporting back to the district. The reports are not publicly available.

#### Data Cleaning: Data Returned from NSC – Match Back

*Initial postsecondary enrollment.* Data returned to AISD from NSC first were linked to the original data file submitted to NSC. Records for enrollment in Summer 2013 only were considered “not matched” to align reporting with best practices.<sup>8</sup> Summer enrollment was defined as enrollment that began on or after June 1, 2013, and ended on or before August 15, 2013.

The following records were considered “not matched,” regardless of postsecondary enrollment:

- Records for which enrollment was withdrawn<sup>5</sup>
- Records for which the student was on a leave of absence<sup>5</sup>
- Records for which the student was deceased<sup>5</sup>
- Records with high school code 991122 (Rosedale High School)
- Records for which enrollment began before:
  - August 15, 2013 (These records represented dual enrollment for the Class of 2013 because they preceded the graduation date; dual enrollment was not an event of interest in the postsecondary annual reports.)
- Records for which enrollment began after:
  - August 4, 2014 (These records represented the second year of enrollment for the Class of 2013; these data were beyond the scope of the postsecondary annual reports because they fell beyond the 2012-2013 academic year.)

Length of enrollment and start/end dates varied across postsecondary institutions (e.g., semesters versus quarters). NSC practices were used, when known, to align our reporting with its automatically generated aggregate reports. The following are NSC-defined variables<sup>9</sup>:

- Fall semester (Semester\_Enrollment = Fall): “High school students who enrolled in a 2- or 4-year postsecondary institution in the fall semester immediately following graduation.”<sup>9</sup>
  - Enrollment that occurs between August 15 and October 31
- Spring enrollment (Semester\_Enrollment = Spring): NSC does not define spring enrollment. However, AISD postsecondary reports included data on spring enrollment. We defined spring enrollment as enrollment that occurs between January 1 and May 31.
- First-year enrollment (Enrolled or Enrolled\_by\_Spring): “The first year after high school

<sup>8</sup> Program-specific reports may use data in which “summer only,” “withdrawn,” “deceased,” and “leave of absence” enrollment are classified as “enrolled.” Reports that use postsecondary data clarify the definition of “enrolled” in the methods section.

<sup>9</sup> Definitions retrieved from NSC *StudentTracker for High Schools - Report Definitions* (as of Fall 2012)

includes any enrollment that occurs between August 15 of the graduation year and August 14 of the following year.”<sup>9</sup>

*Deduplication decision rules.* After data were coded for enrollment, data then were pared down to a single record per student, using the following sequenced selections:

- (1) fall over spring semester,
- (2) 4-year over 2-year institution (“less than 2 year” was considered 2-year),
- (3) full time over part time,
- (4) latest enrollment begin date, and
- (5) latest enrollment end date.

### Reporting Protocols

*Student demographics.* Cleaned data from NSC were linked to demographic data from district student information systems. Demographic data included gender, race/ethnicity, socioeconomic status (defined by free/reduced lunch status), limited English proficiency (LEP) status, special education (SPED) status, and whether or not the student was classified as a career and technical education (CTE) concentrator. A senior is classified as a CTE concentrator if (a) his or her 4-year plan of courses reflected the intent to take a sequence of CTE courses for three or more credits, (b) he or she took one of the CTE courses in the sequence prior to senior year, and (c) he or she took an upper-level CTE course in the sequence during junior or senior year that met the credit requirements. Demographic data were populated using data from PEIMS (100, 101, and 110) and TEAMS (BI, Attendance) tables.

*Scholastic outcomes used to predict postsecondary enrollment.* Academic data were drawn from multiple sources. College readiness status (College\_Ready) was defined by a dichotomous variable (0 = not college ready; 1 = college ready) indicating whether a student met college readiness criteria on at least one mathematics and at least one English language arts (ELA) standardized exam. To be considered college ready, a senior must have met the college ready criteria on the Texas Assessment of Knowledge and Skills (TAKS), SAT, ACT, Compass, or Texas Success Initiative (TSI) test. The criteria for each are as follows:

- ELA
  - Exit-level TAKS:  $\geq 2200$  on ELA test and 3 or higher on essay; or
  - SAT:  $\geq 500$  on critical reading and  $\geq 1070$  total; or
  - ACT:  $\geq 19$  on English and  $\geq 23$  composite; or
  - TSI:  $\geq 351$  on reading and  $\geq 363$  on writing and  $\geq 4$  on essay or  $\geq 351$  on reading and  $\geq 5$  on essay; or
  - Compass:  $\geq 81$  on reading comprehension or  $\geq 6$  on written essay or  $\geq 59$  on writing, if essay = 5.
- Math

- Exit-level TAKS:  $\geq 2200$  on math test; or
- SAT:  $\geq 500$  on math and  $\geq 1070$  total; or
- ACT:  $\geq 19$  on math and  $\geq 23$  composite; or
- TSI:  $\geq 350$  on math; or
- Compass:  $\geq 39$  on algebra

Cut scores that determine college readiness were prescribed by the TEA. Data for college readiness outcomes were limited to valid records (i.e., records in which data were not missing).

The 4-point scale grade point average (GPA) was retrieved from student information systems (from the table known internally as TEAMSBI; variable name: Student\_GPA\_Value). GPAs were ranked into quartiles, based on unweighted 4-point GPA values, using the quartile syntax procedure in SAS-EG. After quartile rankings were computed, a new, dichotomized GPA variable was computed (GPA\_Qrtl), where 0 = not top quartile GPA and 1 = top quartile GPA. The procedure used all graduates as the denominator, and ranked GPAs such that 25% of all records fell into a single quartile.

Attendance during senior year was taken from TEAMS\_Attendance. A new variable for attendance was created via SAS-EG syntax, wherein the total number of days absent (InstDaysAbsent) was subtracted from the total number of days enrolled (InstDaysENR). The new variable was named Days\_Enrolled. All graduates were ranked into quartiles, based on their number of days enrolled. The procedure used all graduates as the denominator, and ranked attendance such that 25% of all records fell into a single quartile. Note that attendance was not used in final reporting because it was not a statistically significant predictor of enrollment when modeled with Class of 2013 data.

*Other predictors of postsecondary enrollment.* Student characteristics used to determine other predictors of postsecondary enrollment were drawn from students' self-report data in the High School Exit Survey (HSES). These characteristics were mother's education level (Mom\_Educ), when the graduate thought about college (Think\_About\_College), and completing at least four college applications (Four\_Apps\_College).

Data on Free Application for Federal Student Aid (FAFSA) (Completed\_Final) application submissions were also included. Student FAFSA completion data were accessed through the federal FAFSA Pilot Project and through the [Texas Higher Education Coordinating Board \(THECB\)](#).

Student characteristics retrieved from the HSES were transformed into dichotomous variables through SAS-EG syntax, and coded as follows:

- Mom\_Educ: 0 = less than college, 1 = college or above
- Think\_About\_College: 0 = during high school, 1 = before high school
- Intend\_Enroll: 0 = did not intend to enroll, 1 = intended to enroll
- Four\_Apps\_College: 0 = less than 4 college applications 1 = greater than or equal to four college applications
- Applied\_Coll\_ANY: 0 = less than 1 college application, 1 = greater than or equal to 1 college application

- Completed\_Final: 0 = did not submit the FAFSA, 1 = submitted the FAFSA

One of the limitations of using students' self-reported data is that it only captures what is occurring at that moment in time. The number of applications submitted, for example, could have increased after the student took the HSES.

*Aspiration gap.* Data linked to student responses in the HSES also were used to determine linkages between intentions to enroll in college (Intend\_Enroll), completing postsecondary education milestones (College\_Ready, Applied\_Coll\_ANY, Completed\_Final), and enrolling in college (NSC data).

Data from the HSES that were used as predictive criteria and used to describe the aspiration gap were limited to students who took the HSES.

*Persistence outcomes.* The database (named PSE\_Class of 2012\_August 2013\_OneRowPerStudent.xls) used to determine persistence outcomes was created in Microsoft Excel in August 2013 by AISD Department of Research and Evaluation (DRE) staff. The database was imported into SAS-EG for further development and analysis.

The following records were considered “not matched,” regardless of postsecondary enrollment:

- Records for which enrollment was withdrawn<sup>5</sup>
- Records for which the student was on a leave of absence<sup>5</sup>
- Records for which the student was deceased<sup>5</sup>
- Records with high school code 991122 (Rosedale High School)
- Records for which enrollment began before August 15, 2012
- Records for which enrollment began after August 14, 2014

Length of enrollment and start/end dates varied across postsecondary institutions (e.g., semesters versus quarters). NSC practices were used, when known, to align our reporting with its automatically generated aggregate reports. The following are NSC-defined variables<sup>10</sup>:

- Fall semester: “High school students who enrolled in 2- or 4-year postsecondary institution in the fall semester immediately following graduation.”<sup>10</sup>
  - Enrollment that occurs between August 15 and October 31
- Spring enrollment: NSC does not define spring enrollment. However, AISD postsecondary reports included data on spring enrollment. We defined spring enrollment as enrollment that occurs between January 1 and May 31.
- First 2 years enrollment: “The first 2 years after high school includes any enrollment that occurs between August 15 of the graduation year and August 14 of the second year.”<sup>10</sup>

The database was then linked to the 2013 NSC return file. The data were then filtered to only include enrollment data for Class of 2012 graduates (WORK.QUERY\_Enrollment\_2012\_Grads).

A new variable, named Persistence, was created to identify enrollment years. Data were coded as

<sup>10</sup> Definitions retrieved from NSC *StudentTracker for High Schools - Report Definitions* (as of Fall 2012)

follows:

- missing = 0 (never enrolled)
- between 15Aug2012 and 14Aug2013 = 1 (Year 1 ONLY)
- between 15Aug2012 and 14Aug2013 = 1000 (Year 2 ONLY)

Columns and rows were transposed so that Persistence values could be summed per record. A new variable, named Persistence\_Outcome, was created by summing Persistence values. Finally, a new variable, named Persistence\_Recoded, was created that recoded the summed Persistence values as never enrolled, year 1, year 2, or persistence. Persistence\_Recoded was coded as follows:

- 0 = never enrolled
- 1-17 = year one only
- 1000, 2000, 3000... 10000 = year two only
- all other values = persisted

After overall persistence rates were computed, students whose Persistence\_Recoded variable equaled “never enrolled” were filtered out of the database before further analysis.

#### Family Educational Rights and Privacy Act (FERPA)

*NSC matching algorithm.* The U.S. Department of Education's Family Policy Compliance Office (FPCO) has certified that NSC practices are [compliant with FERPA](#). Districts share state identification numbers (e.g., social security numbers provided to the district) with NSC, but NSC no longer uses these identifiers for matching purposes (this practice changed in Fall 2008 to comply with FERPA). Current matching logic uses a student’s name, date of birth, and graduation date (last day of attendance).<sup>11</sup> State identifiers are retained in NSC data returned to districts to allow matching back to district records.

*Data suppression protocol.* In our annual postsecondary reports, results were not reported when fewer than five students were in a group, in accordance with FERPA guidelines. In addition, data points greater than 95% were suppressed to mask results that could identify students by omission or by complete inclusion (this issue is often called suppressing “good news”).

*Non-AISD reports on postsecondary enrollment.* At least three annual external report sources provide data on postsecondary enrollment for AISD students. First, NSC automatically generates aggregate reports for the district and for campuses, using the matched data it creates. These reports are returned to the district account along with the student-level data.

The Ray Marshall Center at the University of Texas (UT) at Austin, LBJ School of Public Affairs, prepares an annual report as part of the [Student Futures Project](#). The Ray Marshall Center each year has a data-sharing agreement and research partnership with AISD. As part of this partnership, the DRE sends the Center students’ assessment, graduation, and NSC postsecondary enrollment records. AISD results may differ from results reported by the Ray Marshall Center because the Center reports

<sup>11</sup> The actual match logic used by NSC is proprietary and not publicly available. NSC currently is working on a reference document for end users.

on initial enrollment (fall) only. AISD reports include fall and spring enrollment. Results also may differ due to differences in data-cleaning practices.

The THECB prepares annual reports, referred to as [High School Graduates Enrolled in Texas Higher Education](#). The reports are disaggregated by high school county, school district, and high school in the state of Texas. Note that THECB only reports on enrollment in postsecondary institutions in the state of Texas.

#### Recommendations for Future Work

Predictors of postsecondary enrollment may change over time. Research literature should be reviewed and predictors should be updated accordingly. Also, NSC definitions should be reviewed for updates and changes prior to data cleaning.