

Using NSC StudentTracker for High Schools Reports:

Considerations for Measuring the College Enrollment Rates of High School Graduates

December 19, 2014

Contents

Introduction.....	2
Where do NSC data come from?.....	3
How many college students are covered?.....	3
How many colleges are covered?	3
Which college students are included?.....	4
How StudentTracker works	5
Matching	5
Blocking.....	5
Reporting	6
Putting it All Together: Accuracy of Results	7
Other FAQ's	8

Introduction

Thousands of high schools currently use StudentTracker reports from the National Student Clearinghouse Research Center to measure how many of their graduates go on to college, where, and how many persist to graduation. The reports were designed to help schools to measure their success in preparing students for college, and to evaluate the effectiveness of college access programs and curricula. Increasingly, the results are also being used to inform accountability metrics for schools, districts and states concerned about the performance of secondary schools. In many cases, these metrics are published without fully explaining the source and nature of the data and its limitations.

As the most comprehensive source of national student-level college enrollment data, the Clearinghouse is committed to providing transparency regarding the quality, reliability and accuracy of its data, and of the reports that can be derived from it. Our belief is that the more our data users understand about the data, the better equipped they will be to use it wisely. This document is intended to inform high schools and districts about what to expect from National Student Clearinghouse data accessed through StudentTracker reports.

Where do NSC data come from?

The National Student Clearinghouse Research Center is a 501(c)(3) non-profit organization whose mission is to conduct and support objective research related to educational enrollment, progression and completion for the benefit of students, institutions, and the public. The Research Center provides StudentTracker reports that use data held by the National Student Clearinghouse, a non-profit organization that serves the education community by facilitating the exchange and understanding of student enrollment, performance, and related information.

The Clearinghouse does not survey students. NSC obtains student data directly from registrars, the college and university officers responsible for keeping track of enrollments and degrees. Institutions participate with the Clearinghouse voluntarily. They pay no fee for participating, and they receive valuable services that the Clearinghouse provides based on the data. The institutions typically send to NSC their complete student enrollment list every month, sometimes more often, so the data are very current and constantly updated.

How many college students are covered?

Currently, about 96% of all students who are enrolled in degree-granting institutions in the U.S. are attending schools that regularly send their data to NSC. This percentage has grown substantially over the past decade, but it varies by type of college and by state. It is highest among public institutions (99%), for example, and lowest among for-profit institutions (82%). For-profit institutions typically do not enroll students directly out of high school, however, so their numbers are less relevant to high schools and districts seeking to track their graduates. In fall 2008, for example, less than 1% of traditional-age first-time students entering college in the U.S. did so at a for-profit institution. Excluding these institutions, the current NSC national data coverage rate is 98%.

There are 25 states where the coverage rate for enrolled students, excluding those in the for-profit sector, is 99% or higher, and only seven where the rate is below 97%. The lowest of these (New Mexico) is currently at 95%. Of course, many students enroll in out-of-state colleges, but these numbers provide a good starting point for the maximum percentage of enrolled students that a high school can expect to find in the NSC data.

You can review comprehensive statistics on NSC's data coverage rates by year, college type and state [here](#).

How many colleges are covered?

This is an important distinction. When we say that NSC data covers 98% of the students enrolled in college, that is not the same as saying 98% of the institutions. Since smaller colleges are somewhat less likely to participate in the Clearinghouse than larger ones, our data coverage rate as measured by the number of institutions is lower than the coverage rate measured by the number of students.

Currently, 96% of all public institutions and 73% of all private, non-profit four-year institutions participate in NSC. The denominator for these percentages consists of all degree-granting, Title IV -eligible postsecondary institutions in the U.S. If your high school or district sends a large number of graduates to one or two local institutions that do not participate, this could have a disproportionate effect on the apparent enrollment rates of your students as measured by NSC data.

You can review comprehensive statistics on NSC's data coverage rates by year, college type and state [here](#).

You can find a list of all participating colleges and universities, by name, [here](#).

Which students are included?

NSC data include all students who are enrolled in credit-bearing courses at participating institutions. These include:

- Students who are enrolled full-time and part-time
- Students who are receiving financial aid, and those who are not receiving financial aid.
- Students who attend on campus and on-line.
- Students who are concurrently enrolled in more than one institution, and (in most cases) students who are dual-enrolled in college courses while still in high school.
- Undergraduate and graduate students
- Degree-seeking, certificate-seeking, and non-degree-seeking.
- Students who are enrolled during all terms, including the summer, and those in campus-based study-abroad programs.
- Our data also keeps track of students when they transfer among colleges, even out of state, and when they stop-out and return, even years later.

Which students are not included?

- Students who are enrolled at institutions that do not participate with the Clearinghouse are not included in our data. These include:
 - Most of the US military academies
 - Most tribal colleges
 - Many very small institutions
- International students and undocumented students (non-U.S. citizens) are often not reported to the Clearinghouse, even when they are enrolled at participating U.S. institutions. In those cases when they are reported (we estimate it to be less than half), these students are also more difficult to track if they change institutions.

Limitations in Matching and Reporting

Even though nearly all college students are in the NSC database, they may not all be visible in the StudentTracker report that the NSC Research Center sends to your high school or district. That's because there are limitations in the ability to find, or match, the students in your graduating cohort to those in the NSC dataset. There are also limitations on which student data we are allowed to release once we find it, and limitations that can arise from the different reporting services we provide and how the resulting data can be interpreted by different users. We'll address each of these items in turn in the next section.

How StudentTracker Works

Matching

When a high school, district, or state education authority sends a cohort of graduates to StudentTracker, we attempt to match personally identifying information for each student in the file to the information in the enrollment records in our database. Like any data match, the results depend on the quality of the data provided and the quality of the data stored.

The vast majority of the matches we make, typically over 95% for high school request files, are exact, unique matches based on the student's name and date of birth. For the small portion that requires some additional attention, we may use a number of other factors, including information about the student's high school, graduation date, and enrollment history. Each data element falls within a hierarchy that impacts the overall match confidence. For example, if we are presented with more than one student with the same name and date of birth, we use other elements as additional sources of information to narrow the possible match pool.

Name variations account for the largest number of inexact matches that we encounter, and we have developed robust algorithms to handle most of them with very high levels of confidence. For example, common misspellings, hyphenated names and shortened names are all taken into account in our proprietary logic. Our algorithm allows matches when names differ within a small tolerance level, or by common nicknames and Anglicizations of foreign names, provided that other data elements support the match. This logic incorporates our experience in working with the historical data that we house – well over a billion enrollment records covering 21 years. Our algorithms also accommodate common data entry errors that appear in names and dates of birth (again, provided that they are corroborated by other data elements).

No matching algorithm is perfect, however. State data systems also typically rely on matching to link students in secondary and postsecondary enrollment files, and researchers have recently compared results, for those high school graduates who were enrolled in public colleges, according to NSC's reporting and according to their state's own administrative data. They found that both data sources matched about the same number of students, but not the same students. There were a small number of students, about 2%, that each source found and the other did not (excluding the students who had enrolled at private colleges or out of state, which the state's data did not cover at all). You can find their results [here](#).

Blocking

All NSC services comply with the Family Educational Rights and Privacy Act (FERPA), the federal law governing the protection of student educational data. This law permits schools to release "directory information" for their students to the public. Directory information generally includes name, date of birth and dates of enrollment, and this is what allows us to produce StudentTracker reports. FERPA also gives students over age 18 the right to block the release of their directory information in some circumstances, including the types of research that StudentTracker provides. About 5% of all students exercise this right, which the college registrar records and sends to NSC along with their data. The number varies regionally, however, and there are also a small number of institutions that automatically block all of their students' data from being reported to states. You can find the rates of FERPA blocking, by year, state and institution type, [here](#).

In cases of FERPA blocks, StudentTracker will not identify the student as "found," so someone reading the detailed list of names in a data report might think the student is not enrolled. However, the student will be counted in the

aggregate results that StudentTracker provides, such as the percent of the high school's graduating cohort that is enrolled.

For example, if a high school submits 100 graduates to StudentTracker and 90 of them are actually enrolled (in participating colleges) but 5 of those enrolled students have FERPA blocks on their data, StudentTracker will provide an enrollment rate of 90% in all aggregate reporting for the school. However, the detailed list of names of enrolled students will only have 85. Fifteen students will be shown as "not found" in the detailed list, meaning that they were either not enrolled, or blocked.

This works if the report is generated for a single high school, but if it is generated for an entire state's graduating cohort, the state will only see the single aggregate number for how many students were blocked, and typically have no way to know how many of those belong to each high school. In these cases the schools might end up seeing only the unblocked students in the reports they receive from the state, which would mean that the aggregate enrollment rate for the school in the state-provided report would be lower than the aggregate enrollment rate that the school would see if they had ordered their own StudentTracker report directly from the Research Center.

Reporting

NSCRC offers different StudentTracker reporting services for different types of clients, and in some cases this can affect the results we provide. For example, because state education authorities often provide fewer data elements in their cohort files than do individual high schools and districts, we use different matching algorithms to produce their reports. This can sometimes yield different results for the same students when they are submitted by different requestors.

State clients also often construct their own results, by analyzing the raw data that we provide and calculating individual reports for each school and district. In these cases, calculated outcomes such as "enrolled in the first fall" or "persisted to the second year" can differ in definitions and methods from the ones that NSCRC calculates in the standard reports that we provide directly to high schools and districts. For example, there is no universal definition of "fall term." Colleges and universities use different academic calendars with different term dates, and states may use different cutoffs for fall when looking at the term start dates and term end dates that the colleges provide.

Some high schools also receive results through third-party partners, such as Naviance, that can calculate their own outcomes based on the raw data that NSC provides. Like state reports, these reports are also subject to differences in definitions and methods.

Finally, StudentTracker results depend to some extent on timing. Since NSC data is constantly updated by new college enrollment files, requests that are processed a year, or even a month, apart can yield different results. Sometimes students register late in the term, or withdraw early, and in rare cases a college or university will be late in submitting its datafile to the Clearinghouse. This is why the StudentTracker for High Schools service processes all reports for schools and districts at the same time, three times per year. State reports, however, can be processed at any time, depending on when the state submits the request.

Putting it All Together: Accuracy of Results

Unfortunately there is no single answer to the question of accuracy, since different requestors provide cohort data of different quality (which affects their match rate), for different StudentTracker services (which affects match rate and the handling of FERPA blocks), for students enrolling at different institutions (which affects our data coverage rates). We have worked with individual high schools that independently tracked all of their graduates, through intensive effort via phone calls and personal follow ups, and found their StudentTracker for High Schools results to be as much as 98% accurate. That is, of all the students they independently believed to be enrolled, 98% were actually identified as enrolled in their StudentTracker results. In other words, some schools attain near-perfect matching rates and only miss students who are actually enrolled at non-participating institutions.

Most schools do not experience perfect matching, however. Typographic errors in names and birthdates, name changes, nicknames, etc., all conspire against matching rates, and these can occur in either the high school data or the college data. We estimate that the average match rate for high school cohorts is typically closer to 93% to 95% for the StudentTracker aggregate counts of enrolled students (which include students with FERPA blocks) and 88% to 90% in the detailed student list results (which exclude students with FERPA blocks). That is, the matching process and data errors together reduce schools' ability to find the students in our database by three to five percentage points.

Results that schools receive from state reports that use StudentTracker can be a few percentage points lower still, but it is important to recognize that the state results are designed to compare schools across the state, so as long as all schools are processed in the same way, these comparisons should be "fair."

Some schools and districts can see significantly lower results, however, particularly if they have large numbers of undocumented or international students among their graduates, or if they send a large number of graduates to one or two colleges that do not participate in the Clearinghouse. Our StudentTracker for high schools reports provide a list of the largest in-state institutions that do not participate, to help schools understand these possible variations in local participation rates.

Other FAQ's

1. "I know that a certain student is enrolled in college, but she doesn't show up on my report. Doesn't this prove that the reports are no good?"

These reports are the best available, short of a lot of manual detective-work. Here are some of the most common reasons why a student might not be in your report:

- The student enrolled in a term that does not match the requested dates. (For example, the report is counting enrollments in the "first fall term," but the student didn't enroll until January.)
- The student requested a FERPA block, preventing us from using his or her individual data for research. Nationally, about 5% of all students do this. In these cases, the student would be counted in the aggregate counts that we provide, but would not appear in the detailed list of the students who were counted (by name). If the report was requested by a state or district, they might not be able to tell which of their schools should count the student.
- The student is enrolled in a college that does not participate in the Clearinghouse (about 17% of all public and private non-profit institutions don't participate, but they only enroll about 2% of all college students).
- The student is enrolled in a college that was late in reporting its data to the Clearinghouse. They will show up in a later report.
- The high school or the college data contained an error (e.g. typo, misspelling, garbled birthdate), or the student changed names.
- We could not resolve the match among multiple "identical" candidates (e.g. two students with exact same name, DOB and other data).
- The student is not enrolled in college.

2. "Should we adjust our results to correct for the effects of incomplete data and imperfect matching?"

No data source is perfect. Although we all strive to count every student, another important quality of NSC reports is that they consistently measure outcomes for all schools across a region or across the country. Once each school starts introducing its own adjustments for what it thinks its data should look like, that consistency and comparability of results is lost. On the other hand, if your school is comparing StudentTracker results over time, and you know that one or two institutions that enroll a large number of your students has only recently joined the Clearinghouse, then it may make sense to adjust your older reports for consistency.