Measuring High School to Postsecondary Transitions by State and Year: Opportunities and Challenges

Naihobe Gonzalez, Duncan Chaplin, Brenda Li, and Dan Thal

Transitions from high school to college are challenging for many students, but key for long-term success. Thus, they are a major focus of education policy and research. Key factors that could affect these transitions include K–12 policies and conditions that vary across states and time. However, systematically measuring how students progress from high school to college can be difficult, largely because it requires linking data from disparate K–12 and postsecondary systems.

Currently, there is no central source of data on whether high school graduates are enrolling, persisting, and completing postsecondary education by the state of the high school they attended. Estimates of these numbers are often made available by state and local education agencies—though not all report this information, and those that do use a variety of definitions and are sometimes missing information on transitions to colleges in other states. To be most useful for investigating the impacts of K–12 policies and conditions, transitions measures should include students who attend college in other states and allow comparisons across states and years.

Recognizing the importance of data on educational transitions, the Bill & Melinda Gates Foundation supports multiple efforts to improve data infrastructure, measurement, and reporting across the pipeline from pre-K to postsecondary and beyond. This brief describes an effort by Mathematica, in partnership with the National Student Clearinghouse, to use existing data to produce timely measures of postsecondary transitions by year and state of high school attended. The resulting estimates include college enrollment, persistence, and completion rates for the 2002 to 2019 cohorts of high school graduates in all 50 states. These data, available here, can be used by researchers and policymakers to examine important issues in the transition from high school to college.

Background

Although there is no central source of data tracking the progress of all students from high school to and through college, some data sources have helped researchers explore college enrollment among high school graduates. For example, the Current Population Survey (CPS) sponsored by the U.S. Census Bureau and U.S. Bureau of Labor Statistics collects data on recent high school graduates who enrolled in college as of October of each year. Because the CPS is administered using a probability sample of about 60,000 households, however, the results are measured with substantial error. In addition, the CPS does not collect data on these students' college persistence and completion.

Another source of enrollment data is the Integrated Postsecondary Education Data System (IPEDS), sponsored by the U.S. Department of Education's National Center for Education Statistics (NCES). IPEDS collects aggregate data from every postsecondary institution that participates in

What is the National Student Clearinghouse?

The National Student Clearinghouse was founded in 1993 as a central repository of enrollment data aimed at assisting colleges in meeting reporting requirements from the U.S. Department of Education's Office of Federal Student Aid and other student loan providers.

Although participation is voluntary, an increasing number of colleges have been reporting data to the Clearinghouse over the years. Currently, it receives student records from more than 3,600 colleges representing 99 percent of all postsecondary enrollment in the United States. Enrollment coverage is highest for students in public institutions (99 percent) and lowest for students in for-profit institutions (82 percent).

Colleges report data to the Clearinghouse at least four times per term. Mandatory fields include a student's name, date of birth, permanent address, enrollment status, dates of attendance, and graduation date. Optional fields include race and ethnicity (reported in 59 percent of records), high school code (reported in 19 percent of records), and Pell Grant receipt (reported in 15 percent of records).

In 2004, the Clearinghouse began helping K–12 education agencies track their high school graduates into college. More than 13,000 high schools now submit data to the NSC, representing nearly 40 percent of high school graduates in the United States. The Clearinghouse links students' K–12 and postsecondary records using a matching algorithm and reports data on their college enrollment and completion back to participating high schools, districts, and state departments of education.

Today, the Clearinghouse has become an integral part of the nation's postsecondary data infrastructure. More information about the Clearinghouse's services is available here.◢

federal student financial aid programs. In even-numbered years, colleges report the number of first-time freshmen enrolled in fall who graduated high school in the previous 12 months by the state where they attended high school. Data typically become available with a one-year lag. Using aggregate data on the number of high school graduates each year, it is possible to estimate the percentage of high school graduates in each state who enroll in college in the fall. However, like the CPS, IPEDS does not collect data on college persistence or completion by state of high school attended.

The National Center for Higher Education Management Systems (NCHEMS) has used IPEDS data to report biennial college-going rates of high school graduates by state from 1992 to 2016. Their rates divide the IPEDS college enrollment data by estimates of high school graduates by state, which are based primarily on data from the NCES's Common Core of Data (for public schools) and Private School Survey (for private schools). NCHEMS also estimates persistence and completion rates by state of high school attendance using a formula that implicitly assumes that students attending college in a given state are equally prepared to persist or complete college regardless of where they attended high school.

To produce measures that address the issues discussed above, this project explored using data from the National Student Clearinghouse. Clearinghouse data enable us to estimate transitions with relative precision because they cover almost all college students and not just a sample. Furthermore, Clearinghouse data make it possible to capture the transition from high school to college and beyond because the Clearinghouse has information on students' permanent home address reported when they first enrolled in college. Finally, Clearinghouse data are available just a few months after the start of each term, which means progress can be tracked in a much timelier manner.

Overview

We estimated college success rates for each cohort of high school graduates from 2002 to 2019 based on the state where they attended high school. This section summarizes the data sources and methodology used to develop the measures; for additional information, see the Technical Details section.

Postsecondary success data

The Clearinghouse assigned first-time college enrollees age 19 or younger to their home states based on the permanent home address they first reported to their colleges, allowing us to link postsecondary data to the state where students attended high school. For each entering cohort, the Clearinghouse provided the number of students who enrolled in, persisted in, and completed college, all by home state and year, as described in the box "How did we measure postsecondary transitions?".

To confirm that students' reported permanent home address reflects the location of their high school, the Clearinghouse performed a validation exercise using records from the subset of high schools that participate in the Clearinghouse's StudentTracker for High Schools service.[1] For students age 19 and younger who enrolled in college in 2018, the Clearinghouse compared the state of their high school to the state of the permanent home address they reported to their postsecondary institution that fall. Nationally, at least 97 percent of records had a match. At the state level, the average match rate was also 97 percent, though it ranged from a low of 83 percent for Vermont to a high of 99 percent for Wyoming. Based on these results, we concluded that the

Clearinghouse-reported home state should align closely with the state where students attended high school.

Participation by postsecondary institutions in the Clearinghouse is voluntary and varies across sectors, states, and time. To ensure the measures are as comparable as possible across states and years, we adjusted the enrollment data provided by the Clearinghouse to account for the coverage of enrollment data reported to the Clearinghouse in each state and year, separately for four-year and two-year colleges. We calculated coverage rates by home state using institution-level enrollment counts from IPEDS that report the number of first-time enrollees who graduated high school in the previous 12 months by their state of residence, together with the list of institutions participating in the Clearinghouse and the date when each one began reporting data.

High school graduation data

To obtain an estimate of the annual number of high school graduates in each state, we drew on two sources of data. First, we compiled the number of public high school graduates as reported in the NCES's Digest of Education Statistics. Second, we combined these counts with the NCES's estimates of the number of private school graduates from each state based on the biennial Private School Survey, imputing where counts were missing.[2] Unfortunately, counts of students who complete a high school equivalency diploma by age 19 were not available, even though these students would appear in the postsecondary success data provided by the Clearinghouse if they also enrolled in a postsecondary institution by age 19.

---

[1] For students who entered college in 2018 (the data used for validation), participating high schools represented approximately 40 percent of all high school graduates in the United States: https://nscresearchcenter.org/wp-content/uploads/2018_HSBenchmarksReport_FIN_22OCT18.pdf

[2] It is important to account for private high school graduates because they appear in the postsecondary enrollment counts reported by the NSC. Ten percent of high school graduates in 2016–2017 came from private schools. At the state level, shares of private school graduates ranged from 1 percent (Wyoming) to 25 percent (Hawaii).

How did we measure postsecondary transitions?

To measure transitions from high school to and through college, we created cohorts of students based on the academic year in which they entered any U.S. college and the **state where they attended high school**, which was based on the first permanent home address they reported to their college. To be included in a cohort, students had to enter college by age 19. Data on postsecondary enrollment, persistence, and completion came from the National Student Clearinghouse, and data on public and private high school graduates came from the National Center for Education Statistics. Using these data, we generated the following rates:

Enrollment

Overall college enrollment rate	= # of students enrolled in any college	÷	# of high school graduates in the previous academic year
2-year college enrollment rate	= # of students enrolled in a 2-year college	÷	# of high school graduates in the previous academic year
4-year college enrollment rate	= # of students enrolled in a 4-year college	÷	# of high school graduates in the previous academic year

Persistence

Overall college persistence rate	= # of students who persisted to fall of their second year	÷	# of students in the cohort who enrolled in any college
2-year college persistence rate	= # of students who enrolled in a 2-year college and persisted in any college in fall of their second year	÷	# of students in the cohort who enrolled in a 2-year college
4-year college persistence rate	= # of students who enrolled in a 4-year college and persisted in any college in fall of their second year	÷	# of students in the cohort who enrolled in a 4-year college

Completion

2-year degree completion rate	= # of students who completed a 2-year degree within 3 years	÷	# of students in the cohort who enrolled in any college
4-year degree completion rate	= # of students who completed a 4-year degree within 6 years	÷	# of students in the cohort who enrolled in any college

Benefits and limitations

These measures capture information on transitions from high school to and through postsecondary education for all 50 states from the period of 2002 to 2019, offering several benefits over other publicly available estimates:

- The estimates are based on students' home state and thus can be used to track progress in preparing students for successful postsecondary transitions after high school, including enrollment, persistence, and completion.

- The estimates capture both in-state and out-of-state enrollment. This matters because only 82 percent of high school graduates who go to college do so in state; in states Connecticut, Delaware, New Hampshire, New Jersey, and Vermont, this rate is below 60 percent.[3]

- Data are available annually for 2002 to 2019, allowing for a comparison of longer trends. In addition, the data could be updated within months of the end of each academic year, thus providing more timely information than other data sources like IPEDS.

- By using a single data source for enrollment and for graduation and adjusting for institutional participation in the Clearinghouse, our estimates are comparable across states and years.

- The Clearinghouse captures data on students who enroll at any point over the year, while most other sources report enrollments for the fall only. Reports from the Clearinghouse for 2017–18 show that approximately 22 percent of first-time college students enroll during the spring semester, indicating the importance of capturing data on this group of students.[4]

- By using a flexible age cutoff (that is, students enrolling by age 19) we capture data on some of the students who take a "gap year" before enrolling in college. Many sources only report seamless college enrollment rates, which exclude these students. Estimates from a 2015 study put this figure at about 3 percent of first-time freshmen in four-year colleges.[5] This rate may be even higher overall, including students who enroll in two-year colleges.[6]

At the same time, the estimates have some limitations that are important to consider:

- We cannot account for students who completed a high school equivalency diploma in the denominator used to calculate enrollment rates, even though these students would appear in the count of enrollees if they enrolled in a postsecondary institution by age 19. Nationally we estimate that about 3 percent of students complete high school by age 19 with a high school equivalency diploma,[7] and about a third enroll in a postsecondary institution within one year.[8]

- For recent years, NCES reports estimates of the number of high school graduates because actual data are available with a lag; data on

---

[3] https://nces.ed.gov/programs/digest/d12/tables/dt12_259.asp.
[4] https://nscresearchcenter.org/currenttermenrollmentestimate-spring2018/ and https://nscresearchcenter.org/current-term-enrollment-estimates-fall-2017/
[5] https://heri.ucla.edu/monographs/TheAmericanFreshman2015.pdf
[6] For example, 14 percent of on-time high school graduates in Colorado enrolled in a two-year or four-year college one year after high school graduation: https://cepr.harvard.edu/files/cepr/files/sdp-diagnostic-cg-colorado.pdf
[7] This estimate is based on the number of 16 to 18-year-olds passing high school equivalency tests in 2013, as reported by NCES: https://nces.ed.gov/programs/digest/d15/tables/dt15_219.60.asp.
[8] This means that we may be over-estimating college enrollment for students with regular high school degrees by about 1 percentage point. See https://www.globenewswire.com/news-release/2018/02/21/1372405/0/en/GED-GRADUATES-MAKE-SIGNIFICANT-GAINS-IN-COLLEGE-ENROLLMENT-AND-PERSISTENCE.htm

graduates of private schools are also estimates and are only available every other year. We take averages of adjacent years to impute missing years.

- These rates are estimates and are based on the validated assumption that the first permanent home address students report to their college reflects the state where they attended high school. Although Clearinghouse data suggest the vast majority of students reported the correct state, we do see non-match rates for about 3 percent of the validation sample age 19 or younger, with higher shares in 17 states. Furthermore, we conducted additional validation analyses and found that our measures were consistent with other reported data. See the Technical Details section for more information.

- The cohort definition differs between the numerator and denominator used to measure the initial enrollment rate among high school graduates. The numerator includes students who entered college by age 19 in a given academic year, while the denominator includes all students who graduated from high school in the previous academic year. Very few students complete high school after age 19, so the two definitions should align closely.[9]

- It was not possible to create valid estimates by race and ethnicity, income, or school district due to limitations of the data. See the Technical Details section for more information.

Looking ahead

The estimates produced here allow researchers and policymakers to examine important issues in the transition from high school to college. For example, from 2002 to 2019—as more students graduated from high school—the percentage of high school graduates enrolling in college by age 19 decreased from 80 to 74 percent; persistence and degree completion remained relatively stable during this period. Further analyses could identify bright spot and opportunity states around the country. Analyzing how these rates vary with K-12 state-level policies and conditions could also provide valuable information on the potential of those policies to influence longer-term student outcomes and spur dialogue about possible policy and research implications.[10]

Despite their limitations, these data may be the best source of information on high school to college transitions by state for years to come. As more states develop student longitudinal data systems and additional school districts join the Clearinghouse's StudentTracker for High Schools service, the availability of more detailed data on high school to postsecondary transitions will continue to improve. However, these efforts will probably still be decentralized, limiting the accessibility and comparability of the data. Already, states use different standards for similar transition outcomes. For example, whereas Texas reports in-state college enrollment in the fall after high school graduation, Washington counts

---

[9] However, in any given year some of the students in the numerator (first-time college enrollees age 19 and below) will not be in the denominator (high school graduates from the previous academic year). Those numbers should cancel out across years as long as the age distributions of high school graduates and college entrants remain stable. For example, if 10 percent of students in the numerator in a given year are not in the denominator, we would also expect about 10 percent of the students in the denominator to be in the numerator of a different year.

[10] For example, the rates could be used as outcomes when estimating impacts of high school exit exams (Holme et al 2010; Jacob 2001), rules governing alternative high school credentials (Tyler 2005), mandatory schooling age (Wender 2002), minimum wage laws (Chaplin, Turner, and Pape 2003), age at school entry (Elder and Lubotsky 2009), and general K-12 academic preparation policies (Perna and Titus 2004).

enrollment in any college at any point in the academic year after high school graduation.

In addition to these regional efforts, opportunities exist to build upon the national data infrastructure offered by the Clearinghouse. Recommendations for improving the quality of this data infrastructure were described in a 2016 report by the National Student Clearinghouse Research Center,[11] and included increasing reporting among small postsecondary institutions, improving coverage of existing data elements, adding a small number of key data elements, and developing national standards for defining and measuring priority outcomes. We highlight the following potential priorities for improving the measurement of high school to postsecondary transitions specifically:

- **Require participating colleges to report students' high school code to the Clearinghouse**. This field is already part of the existing data infrastructure, but it is optional; only 19 percent of postsecondary records reported this field in the 2018-2019 academic year.[12] This field would make it possible to report measures of high school to college transitions at different levels, including state, district, and even school. It would also be useful for determining whether a student should be considered an international student.

- **Require participating colleges to report students' race and ethnicity to the Clearinghouse.** This optional field is not reported for a large portion of students, as

noted above. National data from the CPS suggest sizeable gaps persist between the share of White, Black, and Latinx high school graduates that immediately enroll in college, although the survey data are too imprecise in some years to allow for comparisons. Without also collecting information on the high school students attended, however, efforts should be made to align the race and ethnicity classifications used in postsecondary and K-12 data.[13]

- **Develop measurement standards for key outcomes.** Although there is some consensus that initial enrollment, persistence from one year to the next, and degree completion are key outcomes of interest in this field, there are no agreed-upon standards for defining these outcomes. For a long time, states used different definitions to report high school graduation rates, despite this being an important measure of student success. In an effort to standardize measurement practices following the introduction of federal accountability standards that took high school graduation rates into account, in 2005, the National Governors Association introduced the 4-year adjusted cohort graduation rate calculation. The U.S. Department of Education later adopted and mandated this standard.[14] The more visibility that data on high school to postsecondary transitions receive, the greater the incentive will be to develop and adopt measurement standards.

---

[11] https://nscresearchcenter.org/wp-content/uploads/NSC-as-an-Integral-Part-of-the-National-Postsecondary-Data-Infrastructure.pdf

[12] https://nscresearchcenter.org/wp-content/uploads/Reporting_of_NSC_Additional_Data_Elements.pdf

[13] The NSC asks postsecondary institutions to report race and ethnicity using the IPEDS categories, which include a mutually exclusive group for "nonresident aliens" and otherwise differ from the groups used in the NCES data on high school graduates. This misalignment prevents the creation of valid subgroup rates using our approach.

[14] https://www.edweek.org/ew/articles/2008/08/13/45compact.h27.html

Technical Details

In this section, we provide additional details about the data sources and analyses used to develop and validate the estimates of high school to postsecondary transitions.

Data sources

Postsecondary success data

The Clearinghouse assigned first-time college enrollees age 19 or younger to their home states based on the permanent home address they first reported to their colleges. For each entering cohort, the Clearinghouse provided the number of students who enrolled in college at any point during the academic year, the number of that first group who persisted to the fall term of their second year, the number of the first group who completed a two-year degree within three years of entering college, and the number of the first group who completed a four-year degree within six years of entering college. The Clearinghouse provided data for the cohorts entering college in the 2002–2003 through 2019–2020 academic years. Persistence data were available through the 2018–2019 cohort. Two-year degree completion data were available through the 2017–2018 cohort and four-year degree completion data were available through 2013–2014.

The counts include students who reported a U.S. address but who did not have an SSN in the institution-reported data and exclude students identified as being dual enrolled based on their age and high school diploma completion, if this information was available.

High school graduation data

To obtain an estimate of the annual number of high school graduates in each state, we compiled the number of public high school graduates as reported in the NCES's Digest of Education Statistics and combined these counts with the NCES's estimates of the number of private school graduates from each state based on the biennial Private School Survey, imputing where counts were missing. For recent years, NCES has reported projected counts of public high school graduates.[15] When data on private school graduates were missing, whether because data were not collected that year or data were suppressed or flagged as an outlier), we imputed the missing data by averaging the values from the two adjacent years. The last academic year for which NCES reported private high school graduates was 2016–2017. We assumed that the number of private high school graduates in each state remained constant through 2019–2020.

Adjusting for reporting coverage to the Clearinghouse

Participation by postsecondary institutions in the Clearinghouse is voluntary and varies across sectors, states, and time. For example, in 2003 the Clearinghouse captured 87 percent of all postsecondary

---

[15] https://nces.ed.gov/programs/digest/d18/tables/dt18_219.20.asp. NCES data on public high school graduates in 2001–2002 through 2005–2006 came from the "State Nonfiscal Survey of Public Elementary/Secondary Education"; data for 2006–2007 through 2012–2013 came from the "State Dropout and Completion Data File"; and data from 2013–2014 through 2018–2019 came from the State High School Graduates Projection Model.

enrollment nationwide; by 2019, the share had increased to 97 percent (Figure 1).[16] The differences in reporting rates are driven by institutional level and sector. Public institutions (both two-year and four-year) reported more than 99 percent of students in this sector to the Clearinghouse in 2019. In contrast, private two-year colleges had coverage rates below 30 percent. However, these institutions enroll less than 3 percent of first-time students.[17]

Figure 1. Percentage of postsecondary enrollment reported to the Clearinghouse, by institutional level
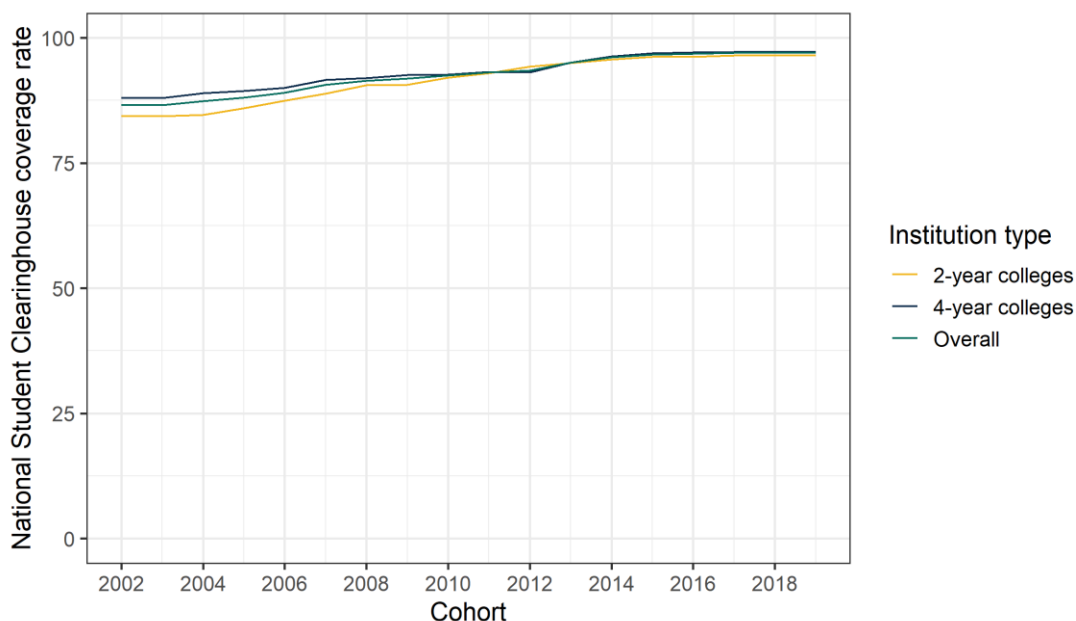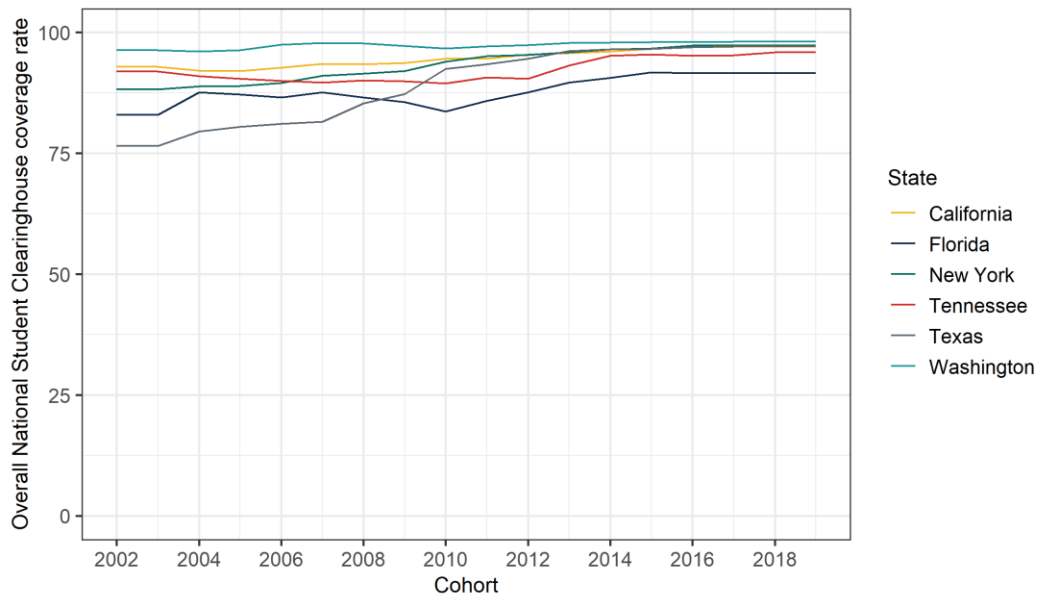

Source: National Student Clearinghouse.

Note: The above coverage rates indicate the percentage of total postsecondary enrollment in the Integrated Postsecondary Education Data System reported to the National Student Clearinghouse.

Coverage rates also vary across states, especially before 2010 (Figure 2; top panel). Although the Clearinghouse's reported coverage rates have increased and converged over time, differences remain. At the state level, coverage rates in 2019 ranged from a low of 91 percent in Colorado to a high of 100 percent in North Dakota and Rhode Island (the state-level average was 98 percent). However, the coverage rates reported by the Clearinghouse are based on the total postsecondary enrollment of the institutions located in each state rather than the first-time enrollment of students who graduated high school in that state.

---

16   https://nscresearchcenter.org/currenttermenrollmentestimate-spring2019/
17   https://nces.ed.gov/programs/digest/d13/tables/dt13_305.10.asp. Note that the share of first-time students in private two-year colleges may be even lower among those who enroll by age 19.

Figure 2. Percentage of postsecondary enrollment reported to the Clearinghouse for select states, by institution state (top) and student home state (bottom)

Source: National Student Clearinghouse.

Note: The above coverage rates indicate the percentage of total postsecondary enrollment in the Integrated Postsecondary Education Data System reported to the National Student Clearinghouse and are based on the location of postsecondary institutions.
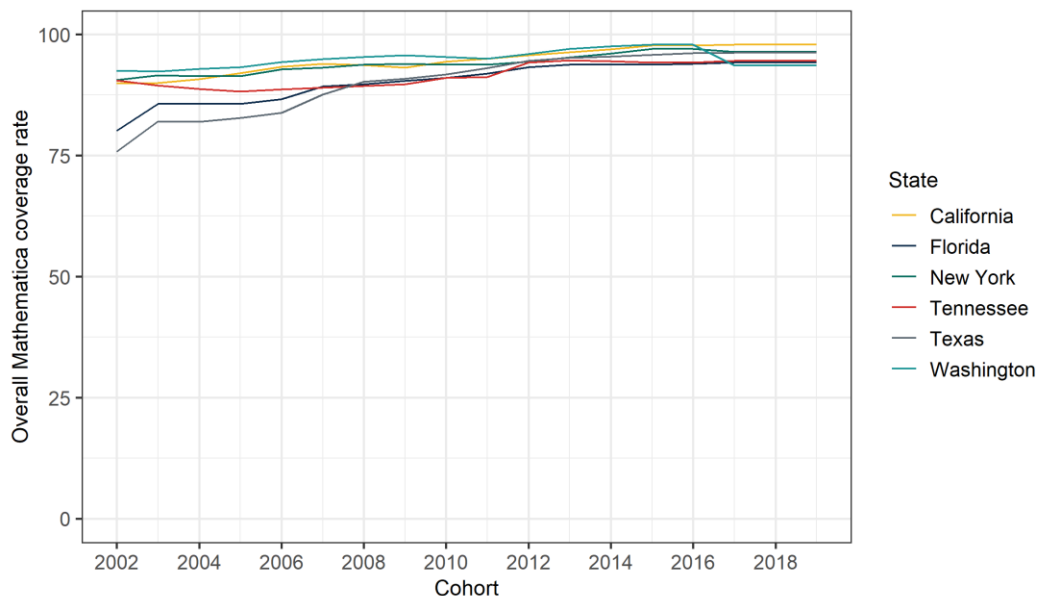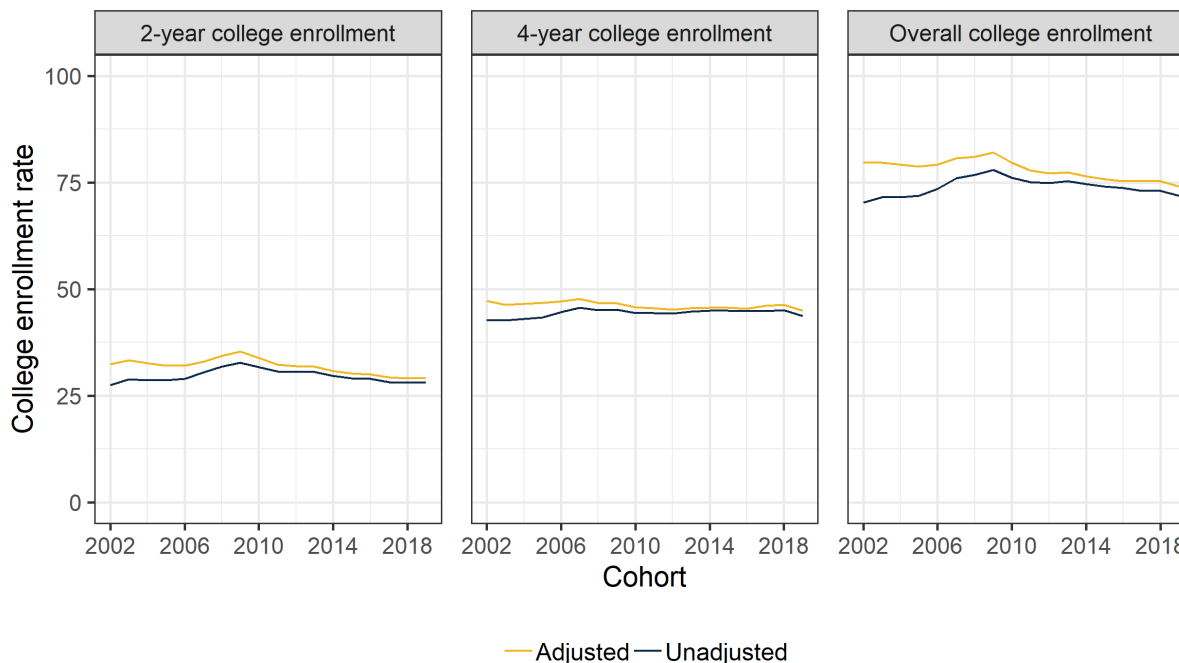

Source: Mathematica calculations, using data from the Integrated Postsecondary Education Data System and National Student Clearinghouse.

Note: The above coverage rates indicate the percentage of first-time postsecondary enrollment in the Integrated Postsecondary Education Data System reported to the National Student Clearinghouse and are based on students' home state.

To account for the varying coverage of students in our data, we calculated our own annual coverage rates using institution-level enrollment counts from IPEDS reporting the number of first-time enrollees in fall who were recent high school graduates by their home state, along with the list of institutions participating in the Clearinghouse and the date when they began reporting .[18] For institutions with multiple branches, we used the earliest participation date in the Clearinghouse as the participation start date for the overall institution. Coverage by home state also increased over time (Figure 2; bottom panel). In 2019, they ranged from a low of 86 percent in Ohio to a high of 99 percent in nine states, including Rhode Island, Massachusetts, and South Dakota. In addition to calculating overall coverage rates, we calculated coverage rates for two-year and four-year institutions separately.

Using these coverage rates, we adjusted the Clearinghouse-reported enrollment counts in each year to account for differences in coverage by both student home state as well as level of institution. Because coverage rates are higher in more recent years, the coverage adjustment has a larger impact on earlier years of data (Figure 3).

Figure 3. Enrollment rates, unadjusted and adjusted for Clearinghouse coverage rates

Source: Mathematica calculations, using data from the National Student Clearinghouse and National Center for Education Statistics.

Note: The reported enrollment rates are calculated by dividing the number of students enrolled in any college by age 19 by the number of high school graduates in the previous academic year. The adjusted rates divide the enrollment count by our estimated coverage rates for each student home state, institution level, and year.

---

[18] Because IPEDS data were only available through 2018 and reporting rates to the NSC have remained consistently high in recent years, we assumed the coverage rates in 2019 were the same as in 2018.

Validating the estimates

To validate our estimates, we conducted several analyses using data from different sources.

Validating the assignment of students' likely home states

To validate whether students' reported addresses reflect the location where they attended high school, the Clearinghouse performed a validation exercise using records from the subset of high schools that participate in the Clearinghouse's StudentTracker for High Schools service. Participating high schools represented approximately 40 percent of high school graduates in the United States in 2018, and were more likely to have high minority enrollments and be in urban locales than high schools nationwide.[19] For students who graduated high school between 2012 and 2018, the Clearinghouse compared the state of their high school to the state in the address they reported to their postsecondary institution in fall 2018.

Nationally, 93 percent of all records had a state match. However, for students age 19 or younger (which our outcome data are restricted to), the match rate was even higher, at 97 percent. At the state level, the average match rate was also 97 percent, though it ranged from a low of 83 percent for Vermont to above 99 percent for Wyoming and West Virginia. Unlike our outcome data, the validation exercise was not restricted to first-time enrollees and thus could include addresses that students reported in a second or other enrollment year. From this validation, we concluded that the state students report at their first postsecondary enrollment by age 19 should align closely with the state where they attended high school.

### Comparing Clearinghouse data against national enrollment counts from the NCES

We compared our national enrollment counts based on coverage-adjusted Clearinghouse data to national data from the NCES, which offers two types of enrollment counts: (1) the total number of first-time, degree-seeking students at any age in the fall, and (2) the total number of first-time, degree-seeking students in the fall who graduated high school in the previous year. Our coverage-adjusted data from the Clearinghouse represents a slightly different student population from either of those counts, as they report the total number of first-time students age 19 or younger who enrolled in fall or spring. For each of these data sources, we compared the enrollment counts for any institution, two-year institutions, and four-year institutions (Figure 4). On average, the adjusted enrollment counts from the Clearinghouse lay in an expected range, above the NCES enrollment counts for recent high school graduates who enroll in fall only and below the NCES enrollment counts for students of any age.[20]

---

[19] https://nscresearchcenter.org/wp-content/uploads/2019_HSBenchmarksReport_FIN_04OCT19.pdf

[20] We intended to also validate our counts against enrollment data from the CPS, which reports the total number of civilian college freshmen age 19 or younger in the fall. However, due to the amount of imputation that would be required for validation due to missing data in earlier years and wide standard errors from the survey sampling, we determined that CPS enrollment counts were not appropriate for this validation exercise.

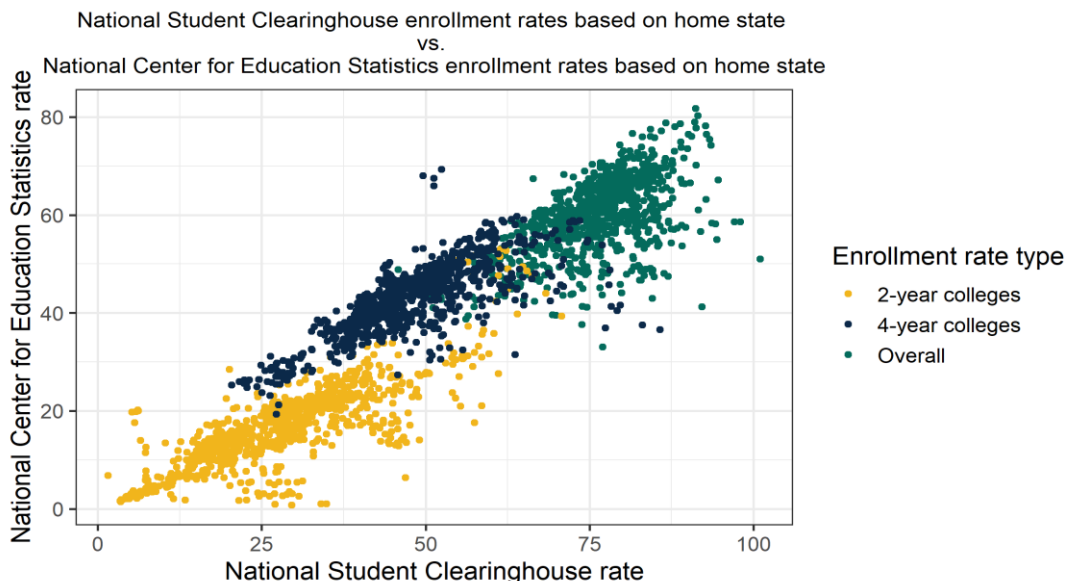**Figure 4. College enrollment counts from the Clearinghouse and NCES**
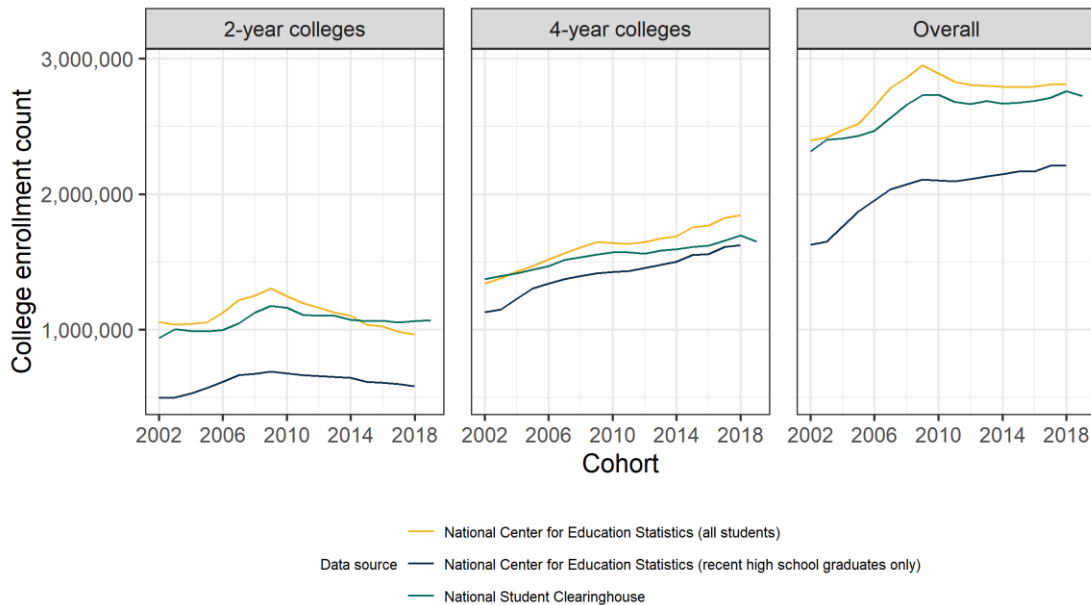
Source: Mathematica calculations, using data from the National Student Clearinghouse and National Center for Education Statistics.

Note: The reported enrollment rates are calculated by dividing the enrollment counts reported by each source by the number of high school graduates in the previous academic year. We divide the enrollment counts reported by the NSC by our estimated coverage rates for each student home state, institution level, and year.

Comparing enrollment rates based on NCES data by students' home state

The rates we calculated were highly correlated with rates based on similar NCES data by home state, as expected. The NCES reports state-level enrollment counts based on student's state of residence. We created state-level rates from these NCES counts using the same denominator (total public and private high school graduates) and compared these rates to our enrollment rates based on Clearinghouse data (Figure 5). We expected a strong, positive correlation between the rates produced by the two sources, but less than 1 because the NCES counts include first-time enrollment in the fall only and across all ages while our counts from the Clearinghouse include first-time enrollment in either the fall or spring for students age 19 or younger. The correlation coefficients between enrollment rates for any institution, two-year institutions, and four-year institutions were 0.56, 0.80, and 0.76 respectively.

Figure 5. Enrollment rates using NSC and NCES data based on student home state

National Student Clearinghouse enrollment rates based on home state
vs.
National Center for Education Statistics enrollment rates based on home state
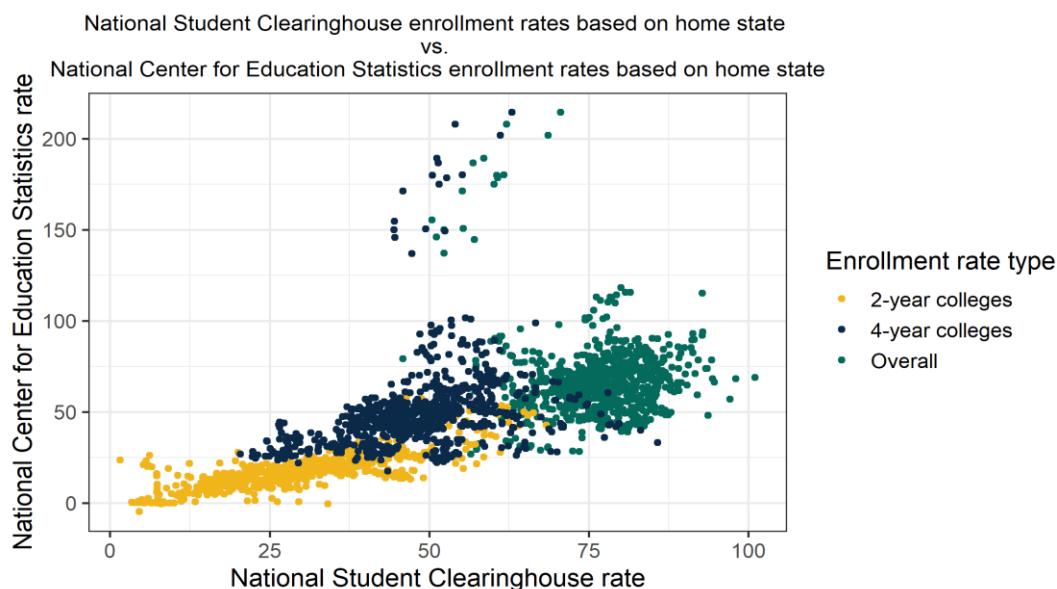
Source: Mathematica calculations, using data from the National Student Clearinghouse and National Center for Education Statistics.

Note: The reported enrollment rates are calculated by dividing the enrollment counts reported by each source by the number of high school graduates in the previous academic year. We divide the enrollment counts reported by the Clearinghouse by our estimated coverage rates for each student home state, institution level, and year.

Comparing enrollment rates based on NCES data by institutional location

As expected, the enrollment rates we calculated using Clearinghouse data had much lower correlations with rates constructed using NCES data based on the state of postsecondary institutions rather than students' state of residence. The NCES reports state-level enrollment counts based on the location of the postsecondary institution. We again created state-level rates from these counts and compared them to our rates based on Clearinghouse data (Figure 6). This check was intended to validate whether the allocation method the Clearinghouse used in assigning students to states indeed reflected their home state rather than where they went to college. Our overall enrollment rates were essentially uncorrelated (correlation = -0.10) with the rates constructed using NCES counts based on the state of the postsecondary institution. Our two-year enrollment rates were strongly correlated (correlation = 0.81), which aligned with our expectations because most two-year students are in-state students. Finally, our four-year enrollment rates were only somewhat correlated (correlation = 0.31) with the rates constructed using NCES counts based on institution.

Figure 6. Enrollment rates using Clearinghouse data based on student home state and NCES data based on institution state

National Student Clearinghouse enrollment rates based on home state
vs.
National Center for Education Statistics enrollment rates based on home state

Source: Mathematica calculations, using data from the National Student Clearinghouse and National Center for Education Statistics.

Note: The reported enrollment rates are calculated by dividing the enrollment counts reported by each source by the number of high school graduates in the previous academic year. We divide the enrollment counts reported by the Clearinghouse by our estimated coverage rates for each student home state, institution level, and year.

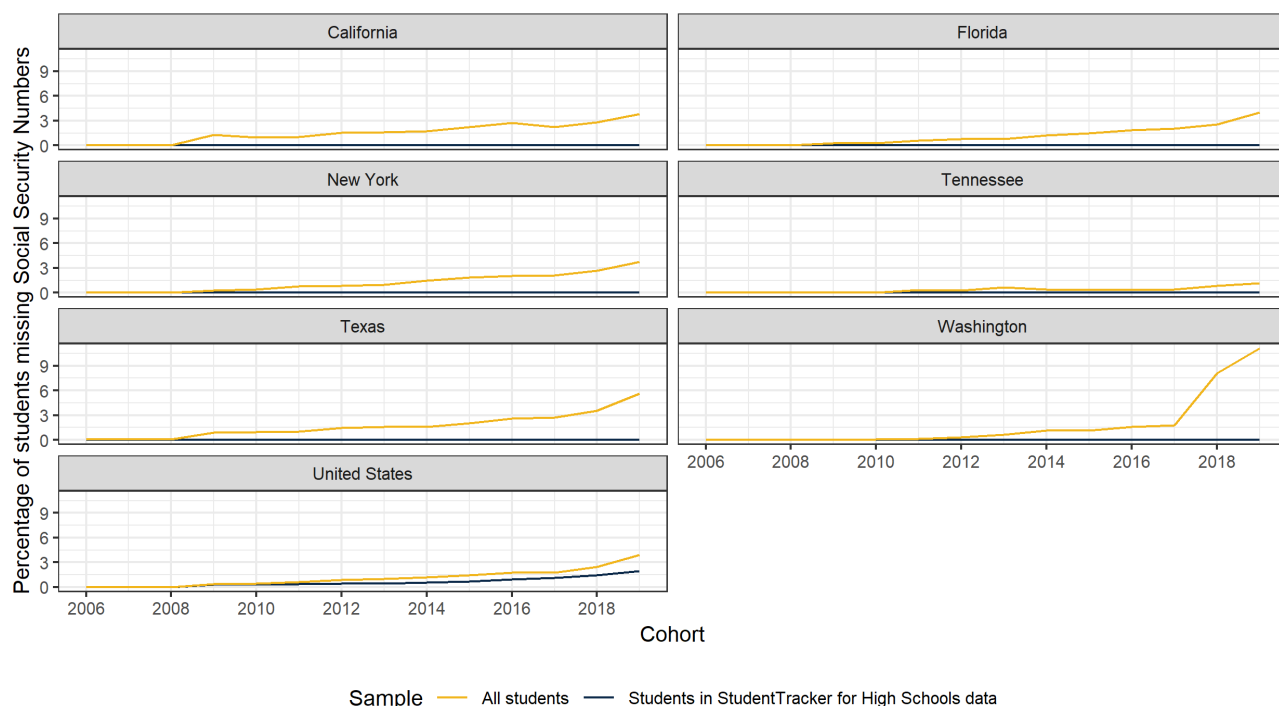## Assessing the inclusion of students without reported SSNs

The enrollment counts reported by the Clearinghouse include students for whom colleges did not report an SSN, which became an optional data field for colleges in 2009. Some of these students could be international students, even though they all reported a permanent home address in the United States. To assess the extent to which this might occur, we asked the Clearinghouse to provide data on SSN reporting rates for all students in our cohort as well as the subset of students with confirmed enrollment in a U.S. high school according to data from the StudentTracker for High Schools service.

Across all states and years, 0.8 percent of students in our data and 0.4 percent of those with a verified U.S. high school did not have a reported SSN (Figure 9).[21] The percentage of students missing SSNs increased each year in both groups, although in recent years the percentage of students missing SSNs increased at a faster rate among all students, reaching about 3 percent in 2019. According to IPEDS data, the percentage of all postsecondary enrollment comprising nonresident aliens increased from 1.2 percent in 2010 to 1.9 percent in 2017, remaining lower than the rate of no-SSN records in the Clearinghouse data. Together, these patterns suggest that SSN reporting rates for U.S. students may have decreased after the field became optional.

---

[21] This group represents a subset of all students who graduated from a U.S. high school, because not all high schools report data to the NSC.

Although the Clearinghouse generally considers students without reported SSNs to be international students, the data are insufficient to make this determination in our case. We opted to keep these students in the counts due to them having a U.S.-based permanent home address and representing an increasing share of students confirmed to have graduated from a U.S. high school.

Figure 9. Percentage of students missing Social Security numbers, United States and selected student home states
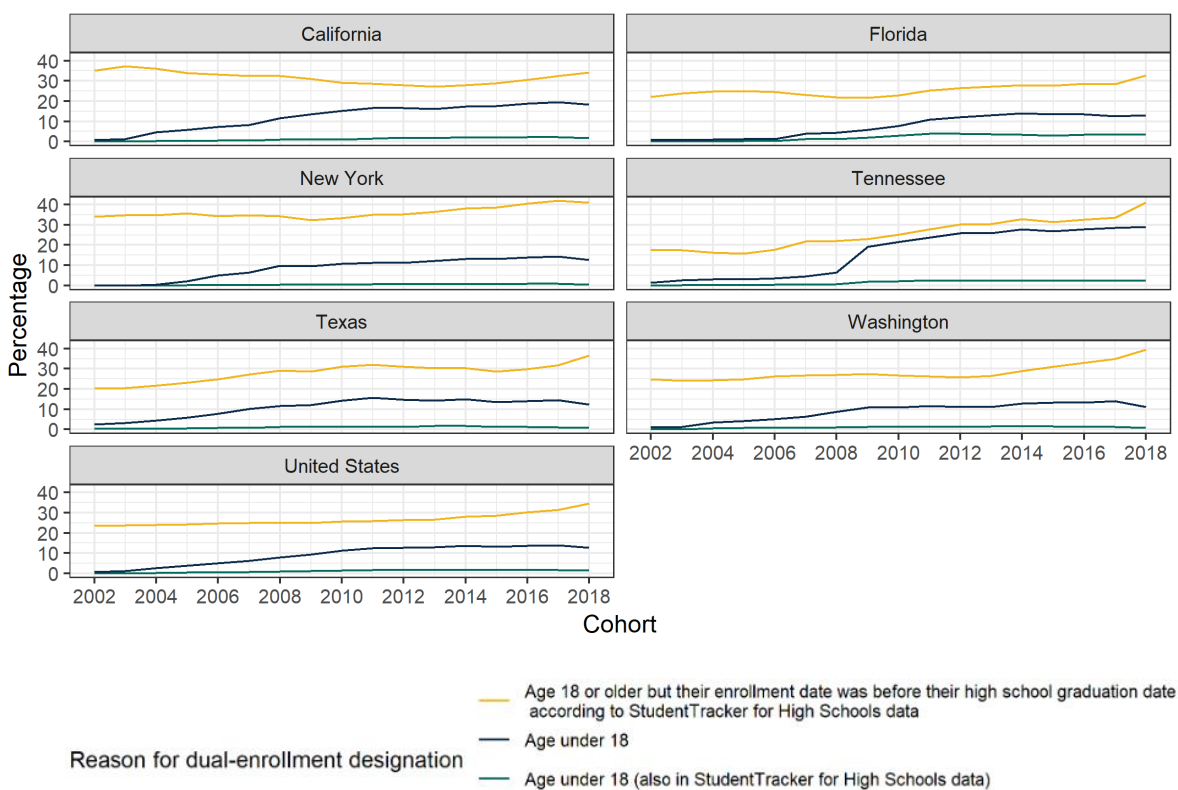

Source: National Student Clearinghouse.

Assessing the exclusion of students inferred to be dual enrolled

The Clearinghouse infers which students are dual enrolled based on their age if the students' high school does not participate in the StudentTracker for High Schools service. Across all states and years, 28.2 percent of students in our data had an earlier recorded postsecondary enrollment before the year counted as their first enrollment (Figure 10). The Clearinghouse determined this earlier enrollment to be a dual enrollment while students were still in high school, either because they were younger than 18 at the time of enrollment or they were 18 or older and their postsecondary enrollment date was before their high school graduation date, if available. This latter determination is only possible for the subset of students appearing in StudentTracker data, however. A potential issue is that StudentTracker reporting rates vary significantly across states and years, which could affect the comparability of the data. However, the vast majority of these students were identified as being dual enrolled due to their age. An average of 88.8 percent of students in each cohort (across all states and years) were younger than 18 at the time of their earlier enrollment. Because dual-enrolled students age 18 or older represent a small share of enrollment, we do not believe this significantly limits the comparability of the date across states and years.

Figure 10. Percentage of students designated as dual enrolled by reason, United States and selected student home states

Reason for dual-enrollment designation

— Age 18 or older but their enrollment date was before their high school graduation date according to StudentTracker for High Schools data

— Age under 18

— Age under 18 (also in StudentTracker for High Schools data)

Source: National Student Clearinghouse.

Measuring outcomes within states

At the start of the project, we explored whether we could use Clearinghouse data to address the lack of transitions data for school districts and key subgroups of students within states. However, there were several challenges associated with developing reliable estimates at these levels.

School districts

If the initial permanent home address students reported to their colleges could be mapped to the district where they attended high school, it would be possible to develop rates at the district level. As an initial test of this assumption, we asked the Clearinghouse to compare the state and city that students reported to their college to the state and city of the high school they attended, for the subset of students who attended a high school that participates in the Clearinghouse's StudentTracker for High Schools service.

Match rates at the city level were much lower than at the state level (62 percent for all students age 19 or younger, and 34 to 86 percent across different states). This could happen due to discrepancies in how students and schools report address information and could thus be improved by using zip codes (which are less likely to have spelling or other differences) to determine the city. In addition, some school districts encompass multiple cities, which means that even if the cities do not match, the student might still have reported the address where they lived when they attended high school. However, we concluded

that without additional analysis and validation we should not use students' first reported city as a proxy for their high school district.

Race and ethnicity

Despite interest in developing rates by race and ethnicity, it was not possible to estimate reliable rates for these subgroups. First, the Clearinghouse had higher than expected rates of missing race and ethnicity data (28 percent of nationwide records in the most recent year). Second, the Clearinghouse asks postsecondary institutions to report race and ethnicity using the categories defined in IPEDS, which include a mutually exclusive group for "nonresident aliens" and otherwise differ from the categories used in the NCES data on high school graduates. Together, these issues resulted in subgroup rates that were often unreasonably small or large (e.g. over 100 percent). The issues were concentrated among the Hispanic, multiracial, and "other" subgroups.

Income

There were also challenges associated with developing rates by income subgroups. The Clearinghouse's income groupings are based on the income level of the census tract of the students' first reported address, which means there is no comparable subgroup definition for high school graduates. [22]

---

[22] As an alternative, we explored creating population-based rates for the income subgroups using estimates of the tract-level population of 17- to 18-year-olds rolled up to the state level. However, these rates frequently exceeded 100 percent for the high-income subgroup, indicating that the alignment between the tract of a student's address and the Census tract population estimates was not good.