



Methodological Notes

Data Coverage

National Student Clearinghouse data track enrollments nationally and are not limited by institutional and state boundaries. As of fall 2024, institutions actively submitting enrollment data to the Clearinghouse account for 97.5 percent of the nation's postsecondary enrollments in degree-granting institutions. The enrollment data used in this report provide unduplicated headcounts for fall starters from 2008 to 2019.

Cohort Identification, Data Cut, and Definitions

This report examines completion and progress towards completion, up to eight years from first enrollment, for cohorts of first time-in-college degree-seeking students who started their postsecondary studies at U.S. colleges and universities in the fall semesters from 2008 to 2019. To limit the cohorts to first-time, undergraduate students only, the study uses data from the Clearinghouse's enrollment reporting and DegreeVerify services to confirm that students included in the study fulfilled the following conditions:

1. Enrolled in a Title IV degree-granting institution in fall semester of corresponding cohort year, excluding U.S. territories (e.g., Guam, Puerto Rico, or the U.S. Virgin Islands).
2. Did not have a previous enrollment record, as shown in the Clearinghouse data, prior to the first day of enrollment in the fall of the corresponding cohort year, unless the previous enrollment record was before the student turned 18 years old (qualified as former dual enrollment students).
3. Did not receive any credential from a postsecondary institution prior to the first day of enrollment in the fall of the corresponding cohort year, according to Clearinghouse data, unless the award date was before the student turned 18 years old (former dual enrollment).
4. Had at least one legitimate enrollment status throughout the study period; that is, enrolled for at least one term with full-time, part-time (i.e., three-quarter-time, half-time or less-than-half-time), or withdrawal status.
5. Showed intent to seek a degree or certificate. That is:
 - For students who started at four-year institutions, enrolled at least one term with an intensity of half-time or higher.
 - For students who started at two-year institutions, either enrolled at three-quarter-time or full-time for at least one term or half-time for any two terms before December 31 one calendar year after the respective cohort year (e.g. December 31, 2020 for the 2019 cohort). We excluded 231,223 students from the 2019 cohort who began at two-year institutions as non-degree-seeking students as a result.

To account for students with postsecondary enrollments due to dual enrollment, students were excluded if they met one or both of these criteria:

1. Earned a high school diploma after the first day of enrollment in the fall of the corresponding cohort year.
2. Were younger than 18 at the time of enrollment.

Race and Ethnicity

Race and ethnicity data are available starting with the fall 2011 cohort. The race and ethnicity categories include Asian, Black, Hispanic, White, Native American, Native Hawaiian or other Pacific Islander, Two or More Races, or Other. The Other category includes Non-Resident (International) and Unknown/Missing. Approximately 14.9% percent of the data are unknown or missing race/ethnicity over the nine cohorts for which these data are available. Unknown and missing rates are higher for the earlier cohorts (22.4% in 2011) but have remained between 11 to 13 percent since the 2015 cohort.

Former Dual Enrollment Students

Our definition of first-time students excludes current dual enrollment students, impacting largely two-year institutions. The cohorts do include *former* dual enrollment students: first-time college students who had previously taken dual enrollment courses. In this report, these are the students who enrolled in college courses while still in high school prior to the fall semester for the corresponding year. Students were identified as former dual enrollment students if they had an enrollment or credential record prior to their diploma data or an enrollment or credential record prior to the corresponding fall semester before they turned 18 years old. Former dual enrollment students represent an average of 21.3% of all the students included in this report. Reference the table below for the average percentage of prior dual enrollment students by sector across all cohorts.

Table 1.1: Former Dual Enrollment Students Across Cohorts by Sector	
Starting Sector	Percentage of Students with Former Dual Enrollment
Public 4-Year	26.4%
Private Nonprofit 4-Year	20.3%
Private For-Profit 4-Year	6.5%
Public 2-Year	17.5%
Public PAB	20.3%

Starting Enrollment Intensity

In this report, starting enrollment intensity is the student's enrollment intensity at the time of postsecondary enrollment. Enrollment intensity is reported as full-time, three-quarter-time, half-

time, or less-than-half-time. Three-quarter-time, half-time and less than half-time students are grouped together and identified as part-time.

Completion and Concurrent Completion

For this report, we examined completion by first-time students. We defined completion as having obtained an undergraduate certificate, associate degree, or bachelor's degree at any institution within the eight-year study period.

Clearinghouse data provide a unique headcount of U.S. college enrollments during each term, which allows for the tracking of individuals including those with concurrent completion. Some students show more than one completion awarded at multiple institutions on the same day. In these instances, the highest credential earned was selected. If the credential type is the same at both institutions, the completion at the starting institution is prioritized.

Definition of Progress and Success

Students are counted in only one outcome category per academic year. Each student's outcome status is based on a hierarchical order of outcomes listed below. Academic years run from July 1st through June 30th.

Graduated – Same School: Student received a credential from the same institution in which they started college. Once a student completes a credential at the starting institution, the student will be counted in this category across all subsequent years.

Graduated – Different School: Student did not receive a credential from the starting institution but did receive a credential from a different institution. Students will be counted in this category in all subsequent years unless they earn a credential at their starting institution in a later year. In that case, the student will be counted as "Graduated – Same School" for that year and all subsequent years.

Enrolled - Same School: Student did not yet receive a credential but is still enrolled at the starting institution.

Enrolled – Different School: Student did not yet receive a credential but is still enrolled at an institution other than the starting institution.

Not Enrolled: Student had no enrollments during the academic year in question and has not earned a credential. This is not a terminal outcome. In other words, a student can be counted as "not enrolled" in one academic year but then counted as "enrolled" in a subsequent year if the student returns to college. Students in this category never enrolled in another institution in any academic year after they stopped attending their starting institution.

Not Enrolled After Transfer: Student had no enrollments for the academic year in question, but in a previous academic year had transferred to an institution other than the starting institution.

Imputation of Values for Gender

The Clearinghouse's coverage of student gender has increased dramatically for enrollments occurring in recent years. However, imputation of gender for the majority of enrollment records is still necessary in order to use the data for research studies using older cohorts. To meet this need, the Research Center developed an imputation process based on first names. Previously submitted name gender pairs throughout the Clearinghouse database are used to determine the probability of any first name being associated with either gender. To increase the accuracy of the imputation process, the Research Center also draws on name-gender data from the Social Security Administration (SSA) and the U.S. Census Bureau. Because the Clearinghouse collects transactional data, its data contain many more unique first names than other sources. The imputation used only those pairs in which the name occurred in at least two instances and was associated with a single gender in at least 95 percent of the instances. The SSA and Census data sets were used to ensure that name-gender pairs were consistent across every data set in which they occurred and to enhance the imputation process by contributing name-gender pairs that did not occur in the Clearinghouse data. The imputation process that yielded additional gender codes produced an average gender coverage rate of 95.4% percent across all cohorts.

Primarily Associate Degree Granting Baccalaureate Institutions (PABs)

As more and more institutions that previously focused solely on granting associate degrees have begun to offer bachelor's degree programs, there has been a surge in IPEDS reclassification of 2-year institutions as 4-year institutions, since IPEDS assigns 2- or 4-year designations based on program offerings. However, many of these reclassified institutions still confer most awards at the associate degree level. These are considered primarily associate degree granting baccalaureate (PAB) institutions.

We utilize the 2021 Carnegie Basic Classification to identify PABs. PABs are defined as institutions that offer at least one baccalaureate degree program and award more than half of their degrees at the associate level. These institutions are made up of two subcategories:

- Baccalaureate/Associate Colleges - Associate Dominant (code 14): institutions that award 90 percent or more of degrees at the associate level, or
- Baccalaureate/Associate Colleges - Mixed Baccalaureate/Associate (code 23): institutions that award more than 50 percent but less than 90 percent of degrees at the associate level.

This method identifies institutions across control groups (e.g., public, private nonprofit) – PAB control is identified using IPEDS data.

Neighborhood Income Measure

The neighborhood income measure provides information about the relative socioeconomic level of students' pre-college neighborhoods for students originating from the 50 U.S. states and Washington, D.C. Neighborhood income is highly correlated with other indicators of neighborhood socioeconomic status (SES) such as home ownership, educational attainment, employment, and poverty. Research suggests that [students hailing from higher SES neighborhoods have better outcomes](#) in terms of college attendance and lifetime earnings, likely due to factors such as [access to high-quality schools, high-achieving peer groups, healthier natural environments, and limited exposure to violence and the criminal justice system](#). It is important to note that this is *not* a measure of family or individual income. Not all students who come from high-income neighborhoods come from high-income families and the same is true of students from low-income neighborhoods.

The measure utilizes street addresses reported to the Clearinghouse with each enrollment submission to locate students' homes in a particular census tract through geocoding. To best approximate socioeconomic origins, we use the first permanent address reported to the Clearinghouse for each student and apply the measure only to students who were 24 or younger at postsecondary entry (non-dual enrolled) and whose first higher education enrollment (including dual enrollment) occurred at age 19 or younger. We apply the measure to cohorts starting in 2010 and on due to the availability of external data sources needed to construct the measure. After these three restrictions—location of the first address, age at first enrollment, and data year of the first address (after 2010)—are taken into account, we are able to include an average of 88.2 percent of students aged 24 and younger across all cohorts for neighborhood income analysis.

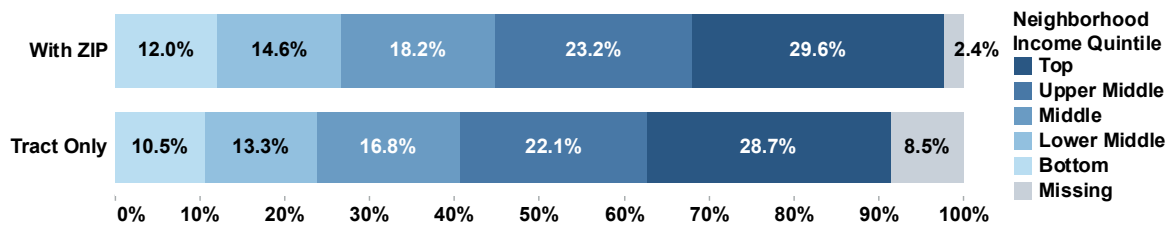
Income data for each tract are sourced from the U.S. Census Bureau's American Community Survey (ACS) five-year estimates. These are adjusted using Regional Price Parity values from the Bureau of Economic Analysis to account for price level differences by state and metropolitan area. The quintiles referenced in this report are of tract median household income adjusted for household size. Quintiles are based on the national distribution of median household income, adjusted for household size, among all census tracts in the 50 states and D.C.

The vast majority (over 90%) of students included in the neighborhood income analyses in this report are successfully geocoded to a census tract. We also include an additional 6.4 to 8.3 percent

(depending on the cohort year) of students who match to ZIP codes but not tracts.¹ ZIP codes are generally larger than tracts, providing less granular measures of a student's pre-college neighborhood. To apply our tract-based neighborhood income measure to these students, we link ZIP codes to Census tracts using crosswalks produced by the US Department of Housing and Urban Development (HUD). A student matching to a ZIP code is assigned weights equal to the share of all residential addresses within the ZIP code lying in tracts of each neighborhood income quintile. For each ZIP-code matched student, the sum of these weights equals 1. Specifically, a student matched to a ZIP code that overlaps with two tracts: one in the lower middle quintile and the other in the middle neighborhood income quintile, with each tract encompassing half the residential addresses in that ZIP code, would be assigned quintile values for each of those quintiles with weights equal to 0.5 for each.

Because our method assumes an equal probability of college-going from tracts of differing neighborhood income levels within the same ZIP code, we are likely slightly overestimating the share of students from lower-income neighborhoods using this method. However, as shown in figure M1, given the relatively small share of students assigned to income quintiles using ZIP code matches, the distribution of students by neighborhood income quintile does not differ greatly between an approach that includes only those students who are geocoded to census tracts and the one used here incorporating ZIP code matches. Inclusion of ZIP code matches also allows us to provide information on neighborhood income background for student groups for whom tract-level geocoding is substantially less successful, such as students from rural areas.

Figure M1. Comparison of Neighborhood Income Quintile Distribution Using ZIP Code Matches and Tract-Only Matches, fall 2019 entering cohort



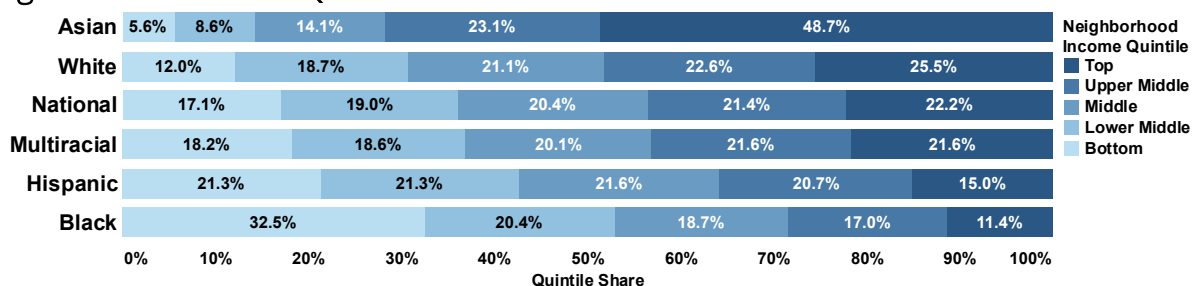
In this report, students with missing neighborhood income quintile values are those who have met the sample restrictions described above, but that (1) we were unable to geocode to either a tract or

¹ An additional small share of students (1.8% to 3.0% depending on the year) match to a ZIP code. However, we exclude these students because their address indicates a PO Box. Since PO Boxes are delivery addresses and not necessarily residential addresses, these students appear as Missing in our neighborhood income analysis.

a ZIP code (including all PO Box addresses) or (2) were geocoded to a tract (or a ZIP code which overlaps such a tract) for which ACS does not publicly publish income data. These missing rates range from 2.4 percent to 3.8 percent depending on the cohort year. The inability to geocode (rather than geocoding to a tract or ZIP code without income data) accounts for nearly all of these missing cases. Geocoding non-matches typically arise from issues relating to the quality of address data such as PO Box addresses and incomplete street addresses. Address quality may be correlated with both neighborhood and household income.

For neighborhood income results by students' race/ethnicity, it is important to note that ethnoracial groups are not uniformly distributed across neighborhood income quintiles. To aid in interpreting these data, we provide a baseline estimate of the population of all 15–17-year-olds in the U.S. residing in tracts of each neighborhood income quintile for each of the race/ethnicity groups we report on, as well as for the Nation overall in figure M2.

Figure M2. National Distribution of 15–17-year-olds by Race/Ethnicity and Neighborhood Income Quintile



Source: U.S. Census Bureau American Community Survey, 2019–2023 5-year Estimates. NOTE: Age 15–17 selected to approximate high school-age and to avoid the undercounting of 18–24-year-olds living away from home while enrolled in postsecondary education. These individuals are not included in home census tract estimates.

Data Limitations

The data limitations in this report center mainly on the data coverage, the methods used for cohort identification, and the definition of key constructs (as outlined above).

Despite the challenges presented by low participation in the early years covered in this report, current Clearinghouse data nevertheless offer near-census national coverage, representing 97.5 percent of U.S. postsecondary enrollments. Keep in mind this report represents the completion rates for Clearinghouse institutions, not all institutions across all cohorts. The Data Coverage tables from our [Working with Our Data website](#) shows coverage by sector depending on the year in question; data should be interpreted accordingly.

Data limitations resulting from the cohort identification methods used in preparing this report should also be noted. Because the Clearinghouse data on designations for class year are incomplete, the researchers identified first-time undergraduate students via two indirect measures:

- No previous college enrollments recorded in the Clearinghouse data, and
- No previous degree awarded in the Clearinghouse's historical DegreeVerify database.

Given these selection criteria, the sample for this report may include students who had enrollments or completions outside of Clearinghouse institutions. It is also possible that a small number of graduate students are included in the study cohort because of inconsistencies in the historical depth of DegreeVerify database records.

Finally, although Clearinghouse data contain some demographic information on students, historical coverage rates for the demographic data elements are uneven. Consequently, results on gender are based partially on imputed values, as described above and an average of 14.9 percent of students over the nine cohorts for which race/ethnicity data are available had the race/ethnicity value unknown or missing. No imputation was attempted for missing race/ethnicity data.

Suggested Citation

Berg, B., Randolph, B., Holsapple, M. & Shapiro, D. (December 2025), Yearly Progress and Completion. Herndon, VA: National Student Clearinghouse Research Center.

The authors thanks Research Center staff members Jeremy Cohen, Jennifer Causey, Ramadan Ibrahim, Hee Sun Kim, and Shannon Lee for their assistance with this publication.