

Data Appendix

Sample and weights

We restrict our sample for each cohort to students that graduated on-time (as defined below) and have available demographic, grade point average (GPA), coursework, test score, and college enrollment data.

For each survey, NCES created several weights appropriate for use with different combinations of variables. For Figure 1, we use the weight for each cohort that NCES recommends when considering postsecondary enrollment outcomes alone. For the remainder of our analysis, because NCES doesn't provide recommended weights based on the exact combination of variables we use, we choose weights that seem the most appropriate for our combination of variables. When there are multiple reasonable weight options, we also consider how the weights would impact our sample size and the results from our analysis. In the end, the weights that we use for the analysis of group differences in enrollment produce slightly higher mean sample enrollment rates compared to the rates shown in Figure 1.

High school completion

For each cohort, we create an indicator for on-time high school completion. This variable is defined slightly differently depending on the cohort, based on available data.

HSLs (2013 Cohort): Students are marked as high school completers if they had a high school diploma, GED, certificate of attendance, or other high school equivalent by November 2013.

ELS (2004 Cohort): Students are marked as having completed high school if they had a high school diploma or GED by August 2004. Students whose exact graduation date was

unknown but had graduated by 2006 are also counted as on-time high school completers (this is a very small share of students).

NELS (1992 Cohort): Students are marked as having completed high school if they had a high school diploma or GED before August of 1992.

HSB (1982 Cohort): Students are marked as completed high school if they had received a high school diploma, GED, or certificate by June 1982.

Demographic groups

Socioeconomic Status (SES)

Our analysis uses each student's socioeconomic status quintile. We assign students to quintiles based on socioeconomic status composite variables that NCES derived using parent/guardian education, occupation, and household income. NCES imputed some of the values when there was missing data. The variables used to create the SES composite were generally collected between students' expected 10th through 12th grade years. **NELS (1992 Cohort)** is a notable exception, using data collected when students were in 8th grade, along with supplemental information from students' later high school years.

Gender

We use biological sex (coded as male or female) as a proxy for gender in all cohorts. We use the composite variable coded by NCES.

Race

We classify respondents in mutually exclusive race/ethnicity categories: Hispanic (of any race), non-Hispanic American Indian or Alaska Native (AIAN), non-Hispanic Asian or Pacific Islander (API), non-Hispanic Black, non-Hispanic white, and non-Hispanic other. For ease of exposition, we refer to race (rather than race/ethnicity) and drop the non-Hispanic modifier.

Race and Hispanic ethnicity were collected in separate questions for all four cohorts. The race categories changed somewhat across surveys. Starting with the **HSLS (2013 Cohort)** cohort, the "Asian and Pacific Islander" category was split to separate "Asian" and "Native Hawaiian or Pacific Islander" categories. Additionally, there is only an explicit "other" category for the **HSB (1982 Cohort)** cohort. For the two later cohorts, respondents could choose more than one race, and respondents who chose multiple races were lumped

together in a “multiracial” category. Our “other” category therefore includes students that identified with two or more race groups for **HSLs (2013 Cohort)** and **ELS (2004 Cohort)**. Because **NELS (1992 Cohort)** does not have a category for “other” or “multiple races,” we do not classify any NELS respondents in our “other” category.

Academic Preparation

GPA

For all cohorts, we use unweighted student GPAs based on only academic courses. In our analysis we divide the GPA variable into deciles within cohort and control for indicator variables for each decile (excluding one group).

Advanced math

We create an indicator for whether students completed any “advanced math” in each cohort. Data for math classes in the two more recent cohorts (**ELS and HSLs**) come from high school transcripts, while data for math classes in the two earlier cohorts (**HSB and NELS**) come from post-secondary transcripts. See below for a list of math classes that we classify as advanced.

HSLs (2013 Cohort): Advanced math includes trigonometry, “other advanced math” (as defined by the HSLs survey), probability and statistics, precalculus, calculus, and any AP/IB math.

ELS (2004 Cohort): Advanced math includes Algebra 3, trigonometry, analytical geometry, linear algebra, mathematics independent study, statistics, probability, introductory analysis, calculus, and any AP/IB math course.

NELS (1992 Cohort) and HSB (1982 Cohort): Advanced math includes calculus, precalculus and trigonometry.

AP/IB courses

We create indicators for whether students took any AP/IB courses. For the two earlier cohorts, we only have information about AP course-taking (not IB). We code AP/IB participation slightly differently across cohorts based on available data.

ELS (2004 Cohort) and HSLs (2013 Cohort): Data come from high school transcripts. We created an indicator variable equal to 1 if students have 1 or more AP/IB credit.

NELS (1992 Cohort): AP participation is determined by a yes/no question asking students in the NELS third follow up survey (two years after high school) if they were ever in the AP program.

HSB (1982 Cohort): We use the H.S. Transcript Study Taxonomy to get information on AP participation. We know how many credits they have taken of each AP class, and if students were enrolled in at least 0.5 credits of at least one AP class then they are counted as having participated in the AP program.

Math and reading tests

Students in each of the cohorts took a standardized math test and in the earlier cohorts took a reading test also. These tests varied by cohort. In our analysis, we use a standardized version of students' scores and assign students to their score decile within cohort. We only use math scores in our regression analysis because they are consistently available across cohorts, but we report both score types in our summary statistics when applicable.

HSLs (2013): We use the HSLs math test administered to students in 11th grade.

ELS (2004 Cohort): We use the ELS math test administered to students in 12th grade.

NELS (1992 Cohort): We use the NELS math and reading tests administered to students in 10th grade.

HSB (1982 Cohort): We use the HSB math and reading tests administered to students in 12th grade. HSB has a two-part math test. In our analysis, we use the data from part 1.

College enrollment

For each cohort, we create indicators for whether a student enrolled in any college (including both two-year and four-year colleges) and whether a student enrolled in a four-year college in the approximately 18 months after their expected high school graduation (June or July of their expected graduation year through December of the following year). Students who enrolled in both a two-year and a four-year institution were coded as enrolling in a four-year. For the two later cohorts (**ELS and HSLs**), the data indicate whether an institution is a two-year or four-year institution (or less than two-year). Classifications for the two earlier cohorts (**HSB and NELS**) did not correspond exactly to those categories.

NELS (1992 Cohort): Students that enrolled in a public two-year or private not-for-profit less than four-year institutions are categorized as having attended a two-year college. The latter category could include private NFP less-than-two-year institutions. Students that attended a four-year public or private not-for-profit are categorized as having attended a four-year institution. All other students are classified as having not attended college.

HSB (1982 Cohort): Students that enrolled in a public two-year or any private less than four-year are categorized as having attended a two-year college. The latter category could include private NFP and private for-profit less-than-two-year institutions. Students that enrolled in a public or private four-year college are counted as having attended a four-year college. All other students are classified as having not attended college.