

# Growth Interrupted:

## An Update on Dual Enrollment Participation Among Public High School Graduates

### Technical Note on Cohort Analysis

October 2024

Kramer A. Dykeman

#### Acknowledgements

This technical note provides detailed illustration of how dual enrollment rates and trends would differ depending on cohort definition. The expertise and experience contained herein is the product of a long-standing partnership between the California Education Lab at University of California, Davis, and the California Department of Education. The partnership is funded by the College Futures Foundation; the Institute of Education Sciences, U.S. Department of Education, through Grant R305A210217 to the Regents of the University of California; and the University of California's Multicampus Research Programs and Initiatives funding. This technical note benefitted from the feedback of multiple colleagues at UC Davis and the CDE including Jonathan Isler and Marvin Cao. The recommendations herein are those of the author alone and do not necessarily reflect positions or policies of funders.

## Technical Note on Cohort Analyses with California Department of Education Data

This technical note documents considerations for defining high school cohorts using data from the California Department of Education (CDE) when investigating student outcomes; the considerations and recommendations are based on the experience of research personnel from the California Education Lab (CA Ed Lab) at the University of California, Davis. In this technical note, we walk through the general principles of defining a cohort, principles that are broadly applicable to a wide array of research questions and student outcomes. To illustrate the nuance of these principles, we use a specific example investigating dual enrollment participation among California high school graduating cohorts.

### General Considerations

#### Defining a Cohort

A cohort is generally defined as a group of students who share some characteristics regarding school attendance. Most commonly, educators reference a cohort when talking about students who are all in the same grade or students who enter or graduate high school in the same year. However, for research a more specific definition of cohort is critically important, as a cohort often serves as the denominator of student outcome measures.

As much of CA Ed Lab's research examines the outcomes of high school graduates, we must frequently consider various ways of defining a *high school graduating cohort*, subject to data constraints. To that end, we have formed six definitions of a high school graduating cohort:

- College and Career Readiness Indicator (CCI) four-year cohort
- Students included in the 11th grade standardized assessment score files
- High school completers in a given year (whether or not they were in 12th grade)
- 12th graders enrolled in a given year with a high school-completer exit code
- 12th graders enrolled on April 15
- 12th graders enrolled on October 1

However, when we investigate factors that may impact high school graduation, we want to observe students *before* they graduate. In these cases, we have employed two definitions of a cohort:

- Four-year graduation cohort as defined for federal accountability purposes
- 9th graders enrolled on the October 1 four years prior to the graduation year that is the object of research

## Deciding Which Cohort Definition to Use

Consistent cohort definitions are especially important when tracking an educational measure over time. If the cohort definition changes, the students being observed may also change. Educational measures are often expressed as rates (i.e., the number of students with a certain characteristic divided by the total number of students in the cohort). If the total number of students in the cohort changes, the rate can change even if there is no change in the number of students with the specific characteristic we are interested in measuring. This leads to misleading results. For example, different census dates (October 1 versus April 15) yield different cohort characteristics, as students enrolled October 1 but no longer enrolled April 15 may be less likely to complete high school or attend college. Clear and transparent cohort definitions also matter for rates that are only calculated for a specific point in time in order to reconcile results that differ between various data sources.

## Specific Case: Dual Enrollment Participation of High School Graduating Cohort

In recent work, we investigated dual enrollment participation rates among a given cohort of California public high school students by merging student-level data from the CDE to student-level data from the California Community Colleges Chancellor's Office (CCCCO). To demonstrate the importance of cohort definition, we calculated the dual enrollment rate for high school graduating cohorts from 2015–16 through 2021–22 using the eight definitions of a high school cohort listed above. In general, the various measures of high school cohorts produce similar results: between 2015–16 and 2019–20 dual enrollment participation increased steadily, but for the two more recent cohorts, participation rates stagnated (Table 1). However, the exact percentage of students participating in dual enrollment depends on the cohort definition. For example, for the 2016–17 cohort, defining a high school cohort by using 12th graders enrolled on April 15 produces a dual enrollment rate of 15.0 percent; while using 9th graders enrolled on October 1 three academic years prior produces a dual enrollment rate of 22.2 percent. Those are substantively different rates, but the only difference in the method of calculation is which students are included in the cohort—that is, the denominator (Table 2).

The first rows in Tables 1 and 2 use a cohort constructed from the CCI, and the second rows use a cohort determined by the four-year high school graduation cohort data. These cohorts are constructed by the CDE for accountability purposes. As accountability rules change over time, different students are included or excluded. For example, between the 2017–18 and 2018–19 cohorts, the students included in the CCI accountability metrics changed.<sup>1</sup> As a result, the number of students in the 2018–19 cohort decreased by about 30,000 compared to the 2017–18 cohort, a much higher number than would be expected from standard enrollment shifts.

---

<sup>1</sup> For more information see the annual Technical Guides for the California School Dashboard and CCI at [www.cde.ca.gov/ta/ac/cm/dashboardccr.asp](http://www.cde.ca.gov/ta/ac/cm/dashboardccr.asp)

Using the CCI cohorts, we observe an increase from 18.0 percent to 21.5 percent in the dual enrollment participation rate during that same time period. However, because of the changes in the CCI cohort, the 3.5 percentage point increase is likely an overestimation of the actual increase of dual enrollment participation. Similarly, between 2020–21 and 2021–22, there was another change in the CCI four-year cohort, causing it to grow by over 100,000 students; we see a corresponding decrease from 24.5 percent to 21.2 percent in dual enrollment participation.

Similar shifts occur in the cohorts constructed from the four-year cohort graduation data (second rows) between 2016–17 and 2018–19. In both these cases, it is difficult to know how much of the decrease in dual enrollment is the result of changes in student behavior and how much is the product of definitional changes in, respectively, the CCI cohort or the four-year graduation cohort. For this reason, when possible, we choose not to use accountability-based cohorts, including the CCI and the four-year graduation cohort for tracking outcomes over time.

The third rows in Tables 1 and 2 define a graduating cohort based on the set of students who took the state standardized 11th grade assessments—the California Assessment of Student Performance and Progress (CAASPP)—the previous year. This cohort definition may be appealing if test scores are an integral part of an analysis. With the adoption of the CAASPP, all students enrolled in 11th grade during the testing window—whether or not they actually took the exam—are included in the CDE’s test score data files.

There are three caveats for using test score data to construct a cohort. First, standardized assessment was paused during COVID-19; there is simply no data for 2020–21 and 2021–22. As such during this two-year period, cohorts cannot be constructed using the test files. Second, testing policies and practices change over time—as they did in the transition from the California Standards Tests to CAASPP—and this can affect which students are included in test score data. Third, 11th graders are neither entering 9<sup>th</sup> graders nor graduating seniors, the two most conceptually common ways of organizing a high school cohort.

In the fourth rows of Tables 1 and 2, the cohort is defined as all students who completed high school in a given year. The CDE creates data files of high school completers for the purpose of matching students to National Student Clearinghouse data and determining 12-month and 16-month college-going rates. The creation of this data set occurs about a year and a half after graduation, delaying when these files become available to researchers.

Rows 5–8 use the CDE’s California Longitudinal Pupil Achievement Data System (CALPADS) to construct cohorts based on cumulative enrollment data files. All students who are enrolled in any California public school during an academic year are included in these data files. Each student record contains the student’s grade level, enrollment dates, and exit code, among other

fields. Limiting the cumulative enrollment file to 12th grade students with an exit code of “completer” creates a cohort of students that is quite similar to the cohort found in the high school-completer file (Rows 4)—and the file is available much earlier. Unfortunately, in 2016–17, there are almost a million missing exit codes, which is three orders of magnitude larger than most other years. Due to this data-quality issue, we determined that high school completers in the cumulative enrollment file should not be used to create a high school cohort for 2016–17.

Further drawing on the cumulative enrollment files, the cohort definition in Rows 6 includes all students enrolled in 12th grade on April 15, late enough in the academic year that students who are enrolled on that date are likely to finish the year and complete high school, while also being early enough that schools have not yet closed for the summer (during which time most students are not enrolled). This produces a set of students who are similar to the high school-completer set, though not quite as similar as the group of students produced using high school exit codes.<sup>2</sup> The advantage of using this cohort definition, compared to the cohort provided in Rows 4, is the absence of the data-quality issues regarding the missing completion code in 2016–17.

Rows 5 and 6 provide information for high school cohorts based on 12th grade students enrolled on October 1, and 9th grade students enrolled three years previously on October 1. October 1 approximates the CDE’s annual census date. This definition results in cohorts of students that are more comparable to the CDE’s publicly reported statistics of the characteristics of students enrolled on the census date. However, it is important to note that there is a large drop in students between the 2016–17 and the 2017–2018 high school cohorts for files based on students enrolled in 9th grade. This is likely the result of data-quality issues—and corrections—during the early rollout of CALPADS.

## Summary and Recommendations

Through a decade-long research partnership with the CDE, researchers at the CA Ed Lab have gained experience with and expertise in constructing cohorts as the basis for measuring a variety of student outcomes and for answering numerous research questions. We have learned that although there are many options for constructing a cohort, it is critical to use definitions that stay consistent over time and do not change due to accountability rules or data-quality issues. For example, we have chosen to base our ongoing dual enrollment research on high school cohorts constructed from the cumulative enrollment data file for 12th grade students enrolled on April 15.

---

<sup>2</sup> Without the use of exit codes, these students are not necessarily high school completers. For example, some 12th graders may continue to a fifth year of high school before earning their diploma or completing high school. Further, all high school completers are not necessarily 12th grade students; some may be adults, and some may be early completers.

**Table 1: Dual Enrollment Rates (percentage)**

	2015–16	2016–17	2017–18	2018–19	2019–20	2020–21	2021–22
CCI	-	15.4	18.0	21.5	24.4	24.5	21.2
Four-Year graduating cohort	12.9	15.0	16.8	19.2	21.9	21.5	21.3
Grade 11 CAASPP	12.9	15.0	17.9	20.6	23.3	-	-
High school completers	14.1	16.4	19.1	21.5	24.4	24.4	-
Enrolled in Grade 12 with completer exit code	14.1	16.2	19.0	21.2	24.3	23.9	23.1
Enrolled in Grade 12 on Apr 15	12.7	15.0	17.7	19.9	22.8	22.4	22.2
Enrolled in Grade 12 on Oct 1	12.9	15.1	17.7	19.9	22.7	22.3	21.9
Enrolled in Grade 9 on Oct 1	-	22.2	18.1	20.4	22.7	22.3	21.9

*Note:* Cells marked “ - ” indicate that a specific cohort calculation is not available in that year due to data constraints.

**Table 2: Number of Students in High School Cohort**

	2015–16	2016–17	2017–18	2018–19	2019–20	2020–21	2021–22
CCI	-	445,436	452,512	422,446	423,381	422,435	524,405
Four-Year graduating cohort	486,126	493,795	532,033	522,736	518,632	526,203	524,405
Grade 11 CAASPP	479,424	487,791	479,182	466,843	472,107	-	-
High school completers	438,346	437,899	446,492	447,722	443,950	443,021	-
Enrolled in Grade 12 with completer exit code	432,745	375,665	445,097	449,028	445,949	451,196	467,378
Enrolled in Grade 12 on Apr 15	452,233	459,675	461,595	462,925	462,158	465,219	464,310
Enrolled in Grade 12 on Oct 1	485,610	484,055	488,242	491,076	487,447	492,356	495,851
Enrolled in Grade 9 on Oct 1	-	590,492	493,469	482,099	485,718	494,059	493,077

*Note:* Cells marked “ - ” indicate that a specific cohort calculation is not available in that year due to data constraints.