Paul Warren

# California's K–12 Test Scores

## What Can the Available Data Tell Us?

## CONTENTS

Technical appendices to
this report are available
on the PPIC website.

California's K–12 system relies on the Smarter Balanced Assessment
Consortium (SBAC) English and mathematics tests to measure student
academic progress and assess school and district performance. This report uses
publicly available data to explore trends in student performance during the first
three years this test has been in place. Key findings include:

- **In the 2016–17 school year, about 45 percent of 3rd grade students
  performed at proficient levels both in mathematics and English.** In
  English, the proportion of students meeting proficiency standards rises after
  3rd grade. By 11th grade, about 60 percent of students tested as proficient.
  By contrast, in mathematics proficiency rates fall as students move forward.
  By 11th grade, only a third score at proficient levels. Achievement levels are
  much lower for students with disabilities, and low-income and English
  Learner (EL) students.

- **Overall, 2016–17 scores changed little from 2015–16.** This is quite different
  than in the previous year, when students made large gains. This pattern is
  consistent across the seven racial and ethnic groups reported by the CDE. The
  previous higher growth rate may have resulted in part from systemic factors,
  such as better understanding of the SBAC tests, continued implementation of
  the standards, and experience with online testing.

- **Scores for low-income students are consistently low across the state,
  while scores of higher-income students vary more widely by region.** The
  lower scores of regions with larger shares of low-income students reflect not
  only the performance of the low-income group, but also relatively lower
  proficiency levels of the higher-income group.

Many things we wanted to learn about student performance could not be
measured using the publicly available test data. In particular, it does not
provide accurate estimates of achievement growth for major student subgroups.
This is a major problem because the SBAC tests were specifically designed to
assess such growth. Public data on school and district performance are also
problematic—so much so that it cannot be used for calculating school and
district gains. The State Board of Education uses similar data to calculate state
accountability ratings for English and mathematics, and our analysis raises
questions about their accuracy.

These issues warrant the state's attention. Fortunately, the State Board of
Education and the California Department of Education (CDE) are exploring
changes in the way accountability measures are calculated. The state should
also reexamine how student mobility affects school and district accountability
data. In addition, CDE should reassess how it releases annual SBAC test data,
with the goal of making it more accessible. To determine how best to meet the
current range of data needs, the department should work with researchers and
policymakers to revamp its data program.

# Introduction

California has nearly completed implementation of its comprehensive K–12 reforms, known as the Local Control Funding Formula (LCFF) Act. The legislation overhauled state funding and governance of school districts. It also called for a new accountability program, which has been designed by the State Board of Education, to monitor the performance of schools and districts. Before LCFF, the state adopted new tests in mathematics and English that have been given annually to students since 2015. The Smarter Balanced Assessment Consortium (SBAC) developed these tests to assess student knowledge and skills in grades 3 through 8 and grade 11.

A few areas of LCFF are still under development. The network designed to assist districts in the improvement process has not been completed. In addition, how to hold alternative schools accountable is still being discussed. And, importantly, sometime in 2018 the State Board plans to change how growth in achievement is measured. Currently, growth is measured simply by comparing average scores of this year's students with last year's—a method used at the school and district levels, and for all subgroups. As an alternative, the Board is considering measuring the gains each student makes on the tests and aggregating that growth for all students in a school, district, or subgroup.

This report represents the third in our series on SBAC test scores in California (see Hill and Ugo 2016 and Ugo and Hill 2017). We have two objectives in writing it. First, we want to use publicly available data to create a baseline for the system's performance and provide context for understanding student progress in the future. Second, we hope to develop a more concise understanding of the growth of student achievement over the three years of testing. The SBAC tests are sophisticated. In particular, expectations for growth implicit in the scoring differ for each grade and subject. By contrast, past state tests, known as the Standardized Assessment and Reporting (STAR) tests, were not designed to measure the growth of achievement from one grade to the next.

This report has three sections. The first briefly describes the SBAC tests and the two ways they report student scores. The second section discusses 2016–17 results for major student groups, including low-income and English Learner (EL) students, and students with disabilities. The third section uses the SBAC data to examine growth in achievement from 2014–15 to 2016–17.

# The SBAC Tests

California's state tests in mathematics and English were developed by the Smarter Balanced Assessment Consortium, one of two federally funded groups that developed tests of the Common Core standards. California adopted the Common Core standards in 2010 and began testing in 2014–15. The 2016–17 state results represent the third year of testing with the SBAC assessments.

The new tests differ from the previous state tests in several important ways. SBAC tests grades 3 through 8 and 11, while the STAR tests covered grades 2 through 11. Students filled in bubbles on STAR paper answer forms, but SBAC is given on computers and the questions change based on each student's performance. Students who incorrectly answer questions get easier questions and those who answer questions correctly are given more difficult ones. This *adaptive* design of SBAC allows a more accurate assessment of each student's knowledge and skills.

Finally, SBAC builds the tests based on a continuous scale so that achievement growth can be measured as students progress from one grade to the next. The STAR tests' design did not permit this type of growth measure. Both STAR and SBAC report scores based on performance levels. STAR had five levels. SBAC uses four. The top two SBAC levels—Met Standard and Exceeded Standard—signal that students are working at a proficient level. The bottom two levels—Below Standard and Nearly Met Standard—indicate students have not reached proficiency.

## How the SBAC Tests Work

SBAC reports student scores in two ways. Student performance is first calculated as a scale score, with points earned by answering questions correctly. The scale scores are then translated into performance levels, which report student scores based on the state's goals for students. Figure 1 shows how the SBAC mathematics test translates scale scores into performance levels. The vertical axis shows the range of scale score points for the mathematics tests. The three solid lines show the minimum score needed to be included in the Nearly Met, Met, or Exceeds Standard performance levels. Scores that fall below the Nearly Met Standard level are included in the Below Standard level.

The maximum and minimum scores on the test are far apart—more than 400 scale points in third grade, growing to almost 600 points in 11th grade. The four performance levels fit within a rather narrow range in the middle of the distribution. In 3rd grade, the low end of the Met Standard level is 55 points above the minimum Nearly Met score. Similarly, the bottom of the Exceeded Standard range is 65 points above the minimum Met Standard score.

In addition, the SBAC scale scores are not linear—that is, students in the lower grades generally are expected to make larger gains than those in the upper grades. As a result, growth in scale scores at different grades are not comparable. In grades 3 and 4, the minimum score for the Met Standard level grows by more than 40 points each year. In grades 5 through 8, minimum scores of the Met Standard level grow between 15 and 24 scale score points.[1] The other performance levels have similar differences by grade.

---

[1] The 11th grade minimum Met Standard score ticks up sharply from the 8th grade score, but three grades separate these groups, so the trend analysis cannot be applied to 11th grade.

FIGURE 1

Performance levels on the SBAC mathematics test are designed to increase each year
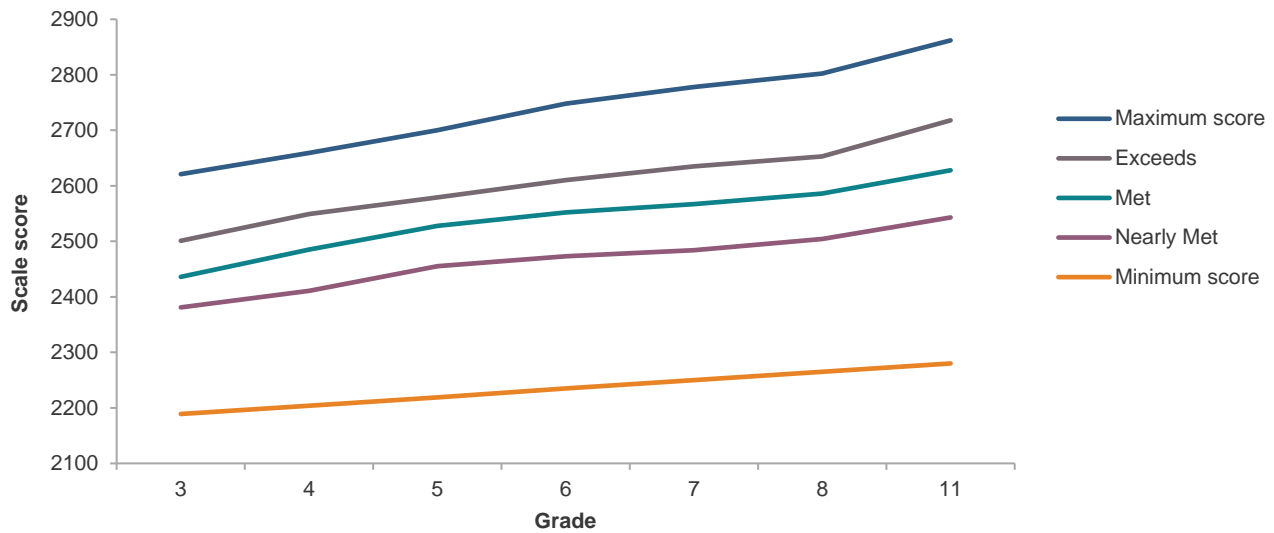


SOURCE: California Department of Education.
NOTE: SBAC = Smarter Balanced Assessment Consortium.

In this report, we use both scale score and performance-level data. When discussing basic performance outcomes, we use the percentage meeting standard or above, which shows the proportion of students meeting the state's learning goals. We use scale scores when examining the growth of performance from one year to the next. This report relies on aggregated data files that the California Department of Education (CDE) posts on its website. These files represent the SBAC performance data available without submitting a specific request that must be approved by the department. The files are extensive, providing information on each school, district, and county, and the state as a whole. Scores for various student groups are also provided, including scores by race/ethnicity, parent education, and language, disability, and income status.

## Test Scores are Imperfect Measures of Performance

The measures used to assess student progress have strengths and weaknesses. Proficiency data, or the proportion of students achieving at or above state standards, are easy to understand, providing a clear sense of how well students are meeting the state's achievement goals. However, these data are not as useful as scale scores as a broad measure of achievement growth.

All group data are influenced by changes in the makeup of the student body from year to year. Change occurs because families relocate or students move into or out of a subgroup. The changes in the composition of a subgroup from one year to the next can lead to erroneous conclusions about its status or growth. Even relatively small changes in the underlying population can affect the meaning of test data. This problem makes it crucial to understand the dynamics of the groups, including the proportion of students moving between schools or between program subgroups, and how those movements affect reported test results.

In addition, test scores are subject to measurement error. Seemingly large changes in aggregate scores in any one year may not represent real increases in what students know and can do. Raising student performance requires changes in the classroom, but these changes rarely generate large increases in student scores. Therefore, at the state level, it is unlikely that growth from year to year will reach a statistically significant level. However, over time, a track record of rising scores can provide more certainty that students are learning more than in the past.
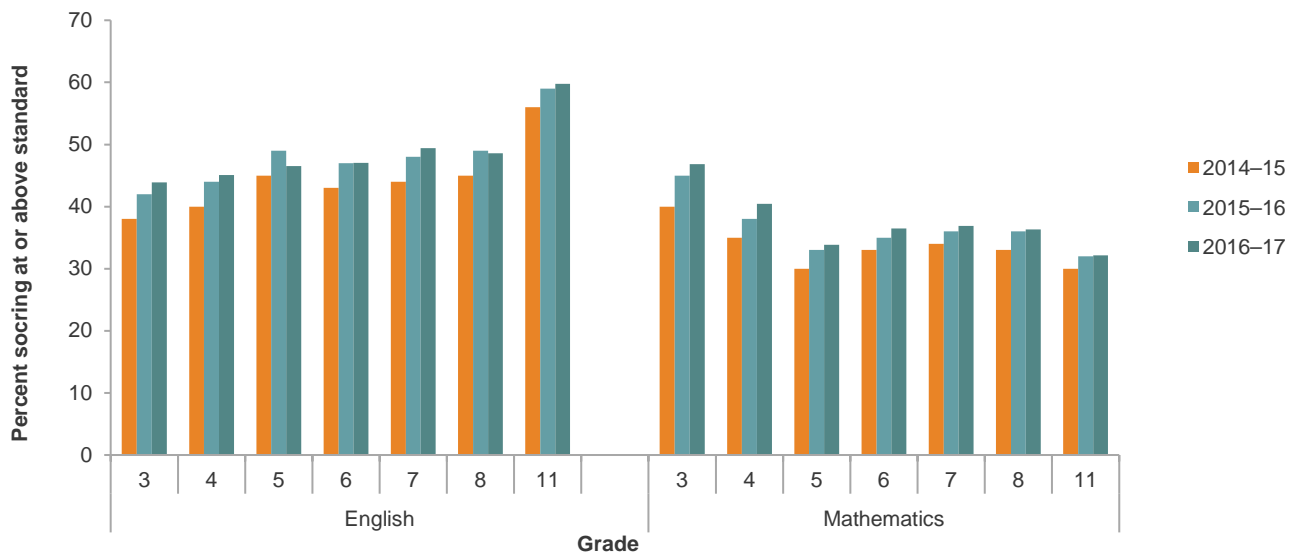
# Percentage of Students Meeting State Standards

In this section, we examine statewide test results in English and mathematics. We mostly focus on results for 2016–17 and leave discussion of progress since 2014–15 until later. We also look at the performance of the major student subgroups, including those defined by race, income, and EL or disability status. In English, achievement generally trends higher as students move up grades. For mathematics, scores are lower in the higher grades. Not surprisingly, student performance in the low-income, EL, and disability subgroups is far below that of the average student in California. However, we find reasons to question whether these data are meaningful in all cases.

## All Students

Figure 2 illustrates SBAC proficiency scores in English and mathematics over the three years of testing. In 2017, almost half of all students scored at a proficient level on the state English examination, that is, at or above the Met Standard level. In grade 3, 44 percent of students met state standard levels. That proportion is higher in the succeeding grades. In grade 11, 60 percent of students were assessed at a proficient level. The figure also illustrates that, in several grades, the proportion of students working at proficient levels increased in 2017. These gains were much smaller and less consistent than the increases registered in 2015–16.

**FIGURE 2**

More students meet state standards on the English test than on the mathematics test



SOURCE: California Department of Education.

NOTE: Percentage of students scoring in the Met Standard or Exceeded Standard level on the state English and mathematics assessments.

Results on the SBAC mathematics test are roughly similar, although performance is lower at higher grades. In 3rd grade, 47 percent of students scored at or above the state standard, about 3 percentage points higher than for the English test. However, that proportion is lower in the higher grades. In 11th grade, only 32 percent met or exceeded the mathematics standard. On average, about 38 percent of students met or exceeded state standards in 2017. In many grades, that proportion increased slightly, but only grades 3 and 4 registered a gain of more than 1 percentage point. And, similar to the English results, gains were much smaller and inconsistent compared with the 2015–16 increases.

California scores somewhat lower than the other 13 states that use the SBAC tests, but made larger increases in 2016–17. The general pattern of performance—rising proportions of students scoring at or above standards on the English test and falling percentages meeting the standards on the mathematics test—is typical of the other states that use the SBAC tests. Fewer students in California scored at the Met Standard level or above on both the English test (5 percentage points lower) and the mathematics test (4 points lower). However, California was the only one of the 14 states to register a gain on the English test in 2017. In mathematics, California had the fourth largest gain of the seven states that reported higher proportions of students working at or above the mathematics standard (McRae 2017).

## Low-Income Students

In this section, we examine the progress of students in 2016–17 based on family income. To simplify this discussion, our analysis focuses on mathematics results. CDE describes this subgroup as "socioeconomically disadvantaged." It includes students from low-income families, foster children, and students from families in which both parents have not earned a high school diploma (CDE 2017a). The vast majority of this group qualify for the federal subsidized meal program intended for students from low-income families.[2] CDE test data do not indicate how many students are added to this subgroup due to the parent education exception, but it is probably small. CDE data show 58 percent of all students qualified in 2016–17 for the federal meals program compared with 62 percent of 2016–17 test takers in the economically disadvantaged group.

The 2016–17 SBAC results reveal large differences between the performance of low-income and higher-income students. This is not a new finding. Indeed, it is a foundation that underlies higher funding levels for low-income students in LCFF. The levels of performance are stark, however. Figure 3 displays 2017 scores for the low-income group by grade. Since low-income students include more than half of all tested students, we use the scores of higher-income students as a comparison. As the figure shows, low-income students perform at proficient levels about 25 percent of the time compared with 54 percent for higher-income students.

**FIGURE 3**
Performance gaps between low-income and higher-income student groups in mathematics are large
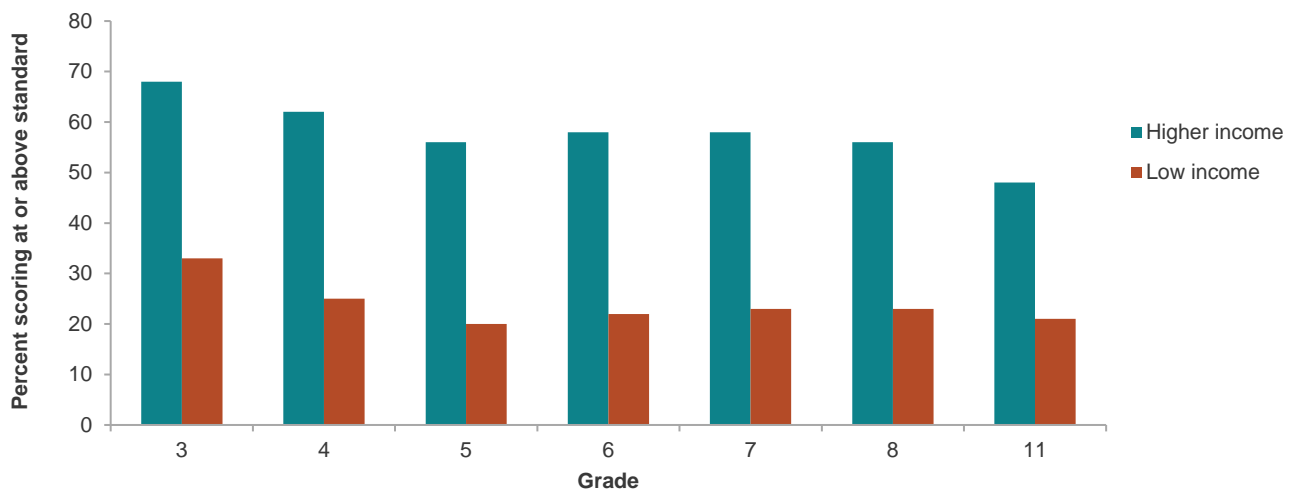


SOURCE: California Department of Education.
NOTE: Percent of students scoring in the Met Standards or Exceeded Standards performance level in 2016-17.

---

[2] Students who come from families that earn less than 185 percent of the federal poverty line, who are foster youth, or qualify for federal migrant education programs. In the United States, 185 percent of poverty equates to about $45,500 in annual income for a family of four. https://www.cde.ca.gov/ls/nu/rs/scales1718.asp.

## Relatively Few Low-Income Students Score at Proficient Levels in All Regions

The performance gap between the two income groups is apparent across the state's regions. At the local level, the proportion of low-income students varies significantly. As a result, school and district test scores with very high proportions of low-income students may look worse than schools and districts with smaller shares of this group simply because of student demographic differences.

Figure 4 shows the influence of income on regional proficiency rates. The figure displays the percentage meeting the standard on the 2016–17 mathematics test in seven regions covering the entire state. The regions are ordered based on the percent of all students scoring at or above the standard. Almost half (48 percent) of students attending schools in the Bay Area achieved proficient scores in mathematics. In the Sacramento metro area and the South Coast (Los Angeles, Orange, and San Diego counties), 40 percent of students met standards. In contrast with these urban regions, California's other regions had fewer students scoring at or above standards. The Central Coast (Ventura to Santa Cruz) had 34 percent of students scoring at proficient levels. North State (all counties north of Sacramento and the Bay Area) had 31 percent, the Inland Empire (Riverside, San Bernardino, and desert counties to the east) had 30 percent, and the South Valley (from San Joaquin to Kern) had 28 percent.

FIGURE 4

Mathematics proficiency rates in 2016–17 vary more for students who are not disadvantaged
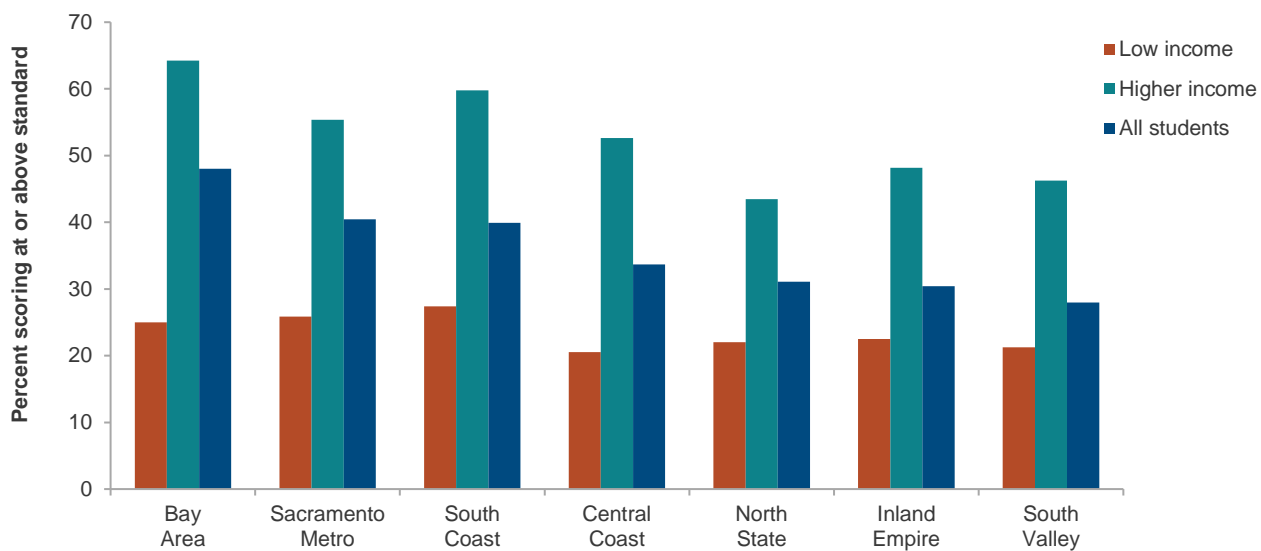


SOURCE: California Department of Education.

NOTE: Percentage of students scoring in the Met Standards or Exceeded Standards performance level in 2016–17.

Figure 4 illustrates several important things. First, the achievement gap is even larger at the regional level than at the state level. In the Bay Area, 64 percent of higher-income students scored at proficient levels while only 25 percent of low-income students did so—a 39 percentage point difference! In the South Coast region, the corresponding figures are 60 percent and 27 percent, a difference of 32 percentage points. And while the gap is smaller in the other regions, it is not because the performance of disadvantaged students was notably higher, but rather the reverse—that the performance of higher-income students was lower.

Second, none of the regions was notably successful in generating high performance from low-income students— only 7 percentage points separate the highest region (South Coast) from the lowest (Central Coast). There is more

variation among regions in the proportion of proficient students from higher-income families, ranging from 64 percent of the higher-income group in the Bay Area to 43 percent in the North State region—a difference of 21 percentage points. However, it is not clear that the higher-income group should be expected to have similar outcomes in all regions. Students in the two income subgroups may be more similar in low-income regions than in higher-income regions. Only 44 percent of Bay Area students are low income, compared with 74 percent in the South Valley.

Third, the figure suggests the All Students average is not a very meaningful measure of a district's output because it hides the divergent scores of the two income groups. The Bay Area earns the highest proficiency rate of all regions (48 percent), 8 percentage points higher than the South Coast region. However, after adjusting for the Bay Area's different student mix, Bay Area schools perform about the same as those in the South Coast region.[3] Similarly, the North State schools report higher total proficiency rates than the Inland Empire, yet both income subgroups perform at higher levels in the Inland Empire. Like the Bay Area, the North State has fewer low-income students than the Inland Empire, which makes its All Students rate higher.

## Race and Ethnicity

Disaggregating scores by race and ethnicity shows large differences in the percentage of students scoring at or above standard. Figure 5 displays the 2016–17 mathematics scores for five race/ethnicity subgroups and the statewide average for all students. Latinos are the largest subgroup, representing 55 percent of test takers in 2017. Only 25 percent of Latinos scored at or above the standard. Whites are the second-largest group, accounting for 24 percent of test takers. Of this group, 53 percent were considered proficient. Asian Americans, representing 9 percent of tested students, were the highest scoring group, with 73 percent scoring at or above standards. African American students made up 5 percent of test takers, with only 19 percent scoring at a proficient level. Finally, the Other category, which includes students identified in one of five smaller subgroups, accounted for 7 percent of students. Only 20 percent of students in the Other category scored at proficient levels in 2017.

**FIGURE 5**
2016–17 mathematics scores differ significantly by race and ethnicity
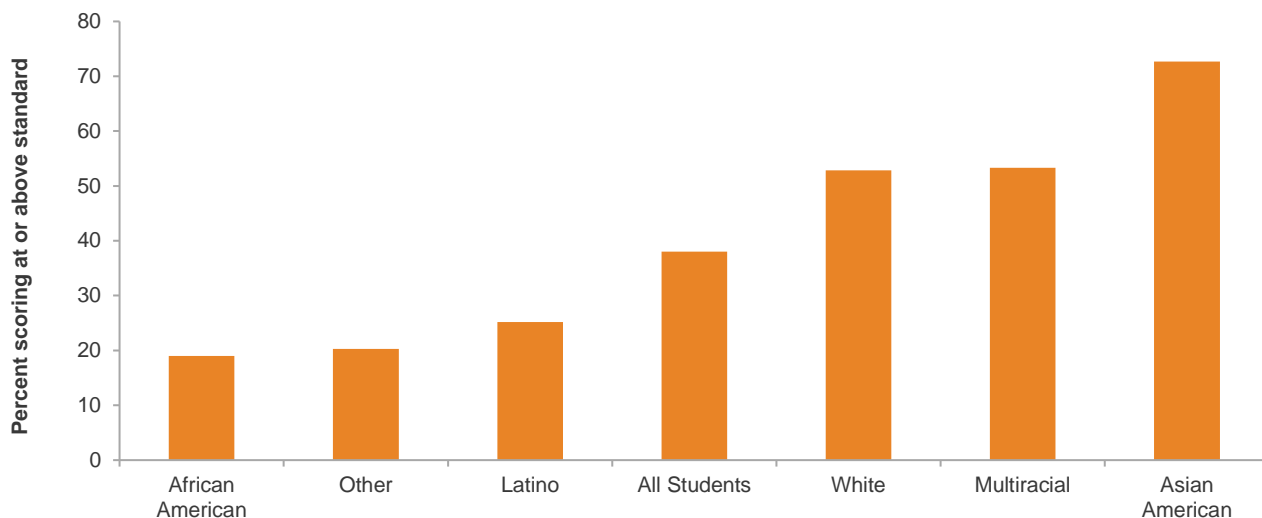


SOURCE: California Department of Education.

NOTE: Percent of students scoring in the Met Standards or Exceeded Standards performance level in 2016-17. Other includes students who identify as American Indian, Filipino, Hawaii and Pacific Islander.

---

[3] This calculation assumes the statewide proportion of economically disadvantaged students.

The achievement divide is highly correlated with family income. Black and Latino students are much more likely to come from low-income families than white and Asian students. Latino students are four times more likely to be from low-income families than from higher-income families (80 percent live in low-income families). Similarly, 74 percent of black students come from low-income families. By contrast, white and Asian American students are much more likely to come from higher-income families (only 28 percent of white students and 35 percent of Asian students were in the low-income group). Thus, much of the differences in proficiency rates by race and ethnicity appear to be driven by family income.

## Students with Disabilities and English Learners

Figure 6 displays 2016–17 proficiency rates of EL students and students with disabilities on the English test. English Learners and former ELs considered fluent in English accounted for 39 percent of students tested. Students with disabilities made up 10.8 percent of test takers. For comparison purposes, the proportion of all students scoring at or above standards is also shown. The levels and trends of proficiency rates for the two subgroups are remarkably similar. In 3rd grade, about 18 percent of students scored at proficient levels. That rate is lower in each subsequent grade. In 8th grade, 11 percent of students with disabilities and 6 percent of EL students performed at state standards. In 11th grade, the proportion of proficient students was slightly higher, but these rates remained far below the almost-60-percent rate for all students.

**FIGURE 6**
Proportion of EL students and students with disabilities performing at standard is low
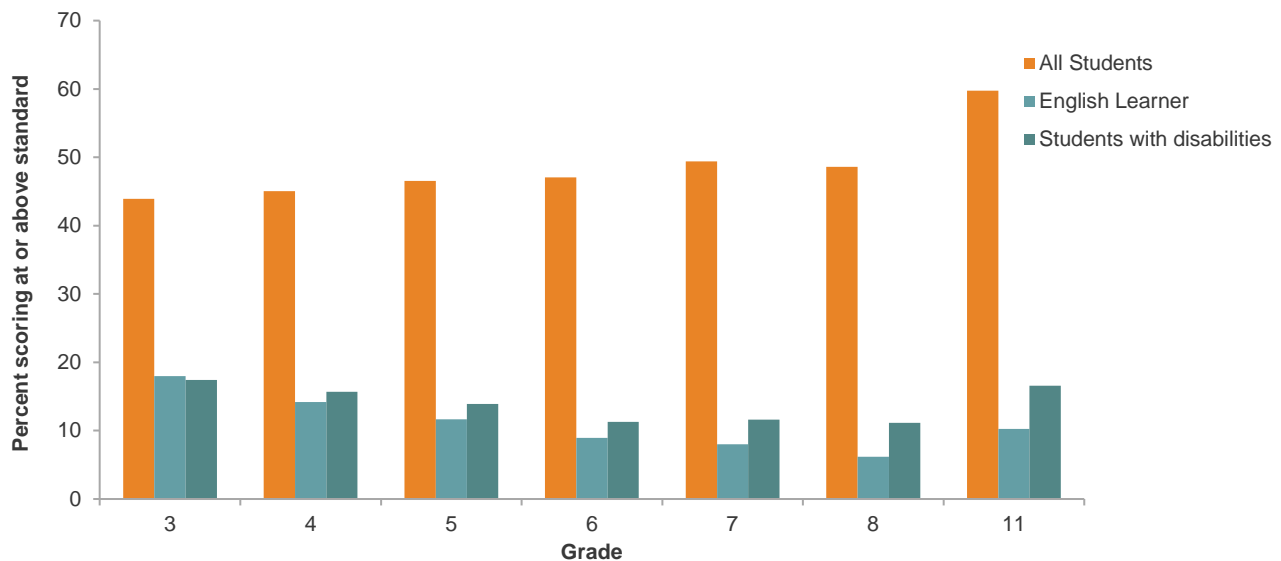


SOURCE: California Department of Education

NOTE: Percentage of students scoring in the Met Standards or Exceeded Standards performance level in 2017. Scores are for the test in English language arts.

However, these data are affected by program dynamics that potentially make student performance for these groups look worse than it actually is. In both programs, new students are identified for services each year and some students "graduate" from the program. For instance, new students arrive in California each year from other

countries and must learn English, while other students improve their English language skills and are reclassified as proficient.[4]

This program migration—the movement of students in and out of programs—makes score trends over the grades difficult to interpret. In 2016–17, the size of the 4th grade special education cohort was 10 percent larger than in the previous grade in 2015–16.[5] The 5th grade cohort was 6.2 percent larger than it had been in 4th grade. However, beginning in 6th grade, the number of students with disabilities actually shrank slightly. With the influx of new students and with some students exiting the program, the number and makeup of students in the program can change significantly.

For the EL subgroup, change is also constant. In each succeeding grade, the size of the EL subgroup is between 15 and 20 percent smaller. For instance, the 4th grade group in 2016–17 was 18 percent smaller than the 3rd grade group the year before. The 5th grade group was 20 percent smaller in 2016–17 than the 4th grade group in 2016. The reduction in students reflects the fact that each year students are reclassified as fluent and no longer needing EL services. The number of newly reclassified students is slightly larger than the reduction in EL students because every year new EL students arrive in California.

Program migration reduces the average scores of students in these programs. The special education and EL programs are intended to help students with special needs succeed in the classroom. By 3rd grade, special education is taking in students who are struggling. And EL programs are adding students who have not mastered English or are new to California. Thus, if we add lower-performing students each year and remove higher-performing ones, group scores will not accurately measure the success of these programs in improving student performance. This may help explain why the scores in Figure 6 decline so much over the grades.

## Ever EL Data Provide a Better Picture

Fortunately, CDE publishes testing data that provide a clearer understanding of EL student progress. The department publishes data on eight student subgroups based on language status.[6] Two of these subgroups—EL and reclassified EL—can be combined to create an "Ever EL" subgroup that avoids the largest problem of EL program migration—reclassification (although, as noted earlier, EL students new to the state continue to affect the averages). Combining the two subgroups allows more accurate tracking of the progress of all students who began as ELs.

Figure 7 shows average 2016–17 percentage proficient for the three EL subgroups by grade. Scores for students who are still English Learners are lowest. By contrast, reclassified EL students score at much higher levels, with 66 percent of 3rd graders and 63 percent of 11th graders scoring at proficient levels. The Ever EL group represents a weighted average of these two subgroups. In 3rd grade, the percentage scoring at or above standards is a low 36 percent. This reflects the fact that 72 percent of 3rd grade Ever ELs are in the EL subgroup. By 11th grade, 75 percent of Ever ELs are in the reclassified EL subgroup, and that subgroup's higher scores boost the average proficiency rate of the Ever EL group.

---

[4] One expert we talked with described the EL population as being like a half-full sink with the tap running and the drain open. The sink never rises or falls, but water is being added and drained at a considerable rate.

[5] The actual change in the number of students entering and exiting the program is almost certainly larger than what the group data indicate because CDE data contain only the net changes in subgroup sizes.

[6] These include students who are English only, reclassified as fluent, were initially assessed as fluent, are currently EL students, are EL students enrolled in a school in the United States less than 12 months, are EL students enrolled in a school in the United States for more than 12 months, and a subgroup that combines English only, reclassified and initially fluent students.

FIGURE 7
The Ever EL subgroup better illustrates the progress EL students made in 2017
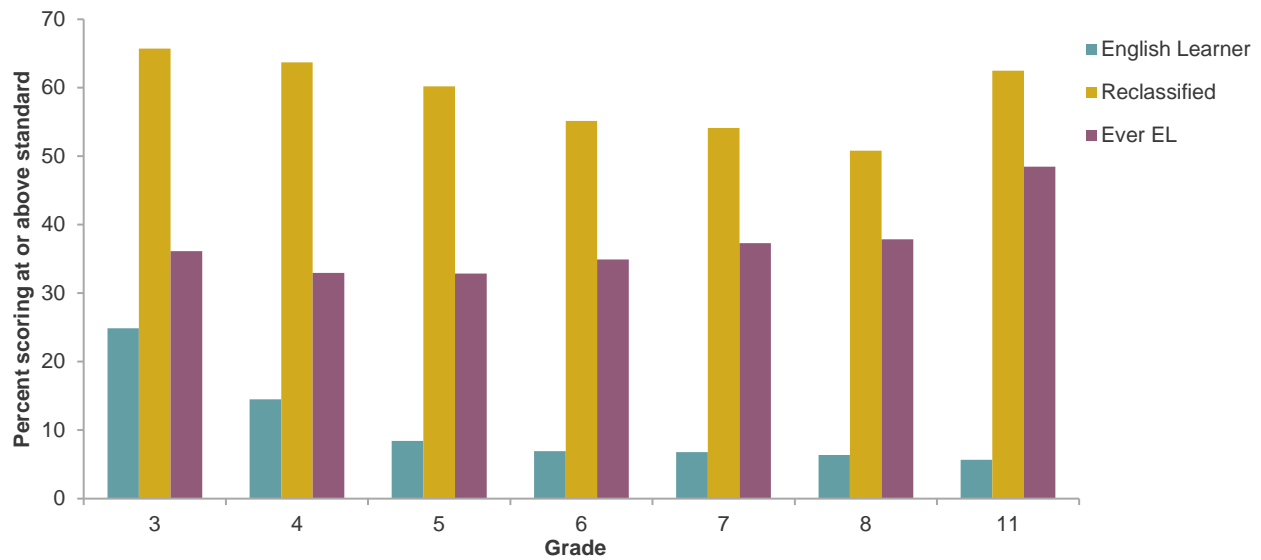


SOURCE: California Department of Education and author's calculations.
NOTE: Ever EL represents the weighted average of the EL and Reclassified EL subgroups.

## EL and Special Education Group Scores Must Be Interpreted with Caution

The data reported for ELs and students with disabilities must be understood in context. The 3rd grade scores offer an accurate snapshot of student achievement, but they mean something different than the 6th grade or 8th grade scores because the composition of these groups has changed. And, as our EL analysis illustrates, the snapshot may understate the progress students in these groups make. In fact, the data show that EL and former EL students account for about 80 percent of the increase in the number of students statewide scoring at or above standards in grades 5 through 8. This degree of success is simply not apparent from the EL data. These issues make using CDE's public use data for these two groups problematic. When using the group data for accountability purposes, the department and the State Board of Education proposed to address these problems by keeping students in the subgroups up to four years after they graduate from the programs.[7] This reduces the distortions created when higher-performing students exit the programs.

# Data Reveal Many Challenges

Our analysis of 2016–17 SBAC scores shows the challenges facing California K–12 schools. In English, about 43 percent of 3rd graders perform at grade level. Over the grades, performance rises, so that six of ten are proficient in 11th grade. In part, this improvement reflects EL student gains. These findings seem to be positive signs for the system.

The picture of mathematics achievement is less encouraging. Mathematics achievement starts relatively strong, with 47 percent of 3rd graders working at standard. But the proportion of students working at or above grade level falls in the higher grades. In 11th grade, only one-third perform at proficient levels. These data suggest that students are not keeping pace with the standards.

---

[7] Specifically, individual-level data are combined to create subgroups that include former EL students for four years and students who were no longer considered disabled for two years (CDE 2017b). The State Board is applying for a federal waiver to use this broader EL group to calculate school and district progress for the purposes of the federal accountability program.

We also found that many large student subgroups perform at lower levels than the statewide average. Members of the largest subgroup—low-income students—are half as likely to score at or above standard in mathematics compared with students from higher-income families. This gap starts in third grade and grows in the higher grades. Large differences in performance are also evident along racial and ethnic lines. However, much of the racial gap reflects income gaps that also fall along racial and ethnic lines. African American and Latino students have proficiency rates half those of white students, but they are much more likely to come from low-income families. This alignment of income and proficiency underscores the importance to our state of the LCFF investment in addressing the achievement gap.

Our examination of scores for English Learners and students with disabilities shows that the data CDE posts in its public use files must be interpreted carefully. The data are accurate—but only as a snapshot of the scores of students who were in the programs in a particular year. It is extremely easy to misinterpret what the data mean. For those who know little about the dynamics of EL and special education programs, it is easy to conclude that the programs do little to improve student performance. The Ever EL group, which CDE has added to its public data release, shows such a conclusion is incorrect.

CDE attempts to rectify the problems of the public use data on EL by publishing the Ever EL data. This provides very useful information about the larger group of students that includes all those ever classified as EL, but it does not tell much about the current group of EL students. The same problem affects special education data. While there may be no perfect solution to the problems created by program migration, our EL analysis shows that some alternatives provide more accurate information. At a minimum, something similar should be created for special education. While CDE publishes data on eight language subgroups, it publishes only two groups for special education—students with disabilities and students with no disabilities. In the next section, in which we examine achievement growth from 2014–15 to 2016–17, program migration surfaces as an even larger problem.

# Growth in Average Scores, 2015–16 to 2017–18

The SBAC tests were designed to measure growth in scores from one grade to the next. Given the large differences among the subgroups in the percentage that score at or above standard, growth becomes a critical indicator of whether students who have not reached proficient levels are catching up.

Our ability to accurately capture growth using public SBAC data files is extremely limited, however. Small differences in the underlying group of students in the subgroups from one year to the next can make the data unreliable. Since every group of students—even the statewide total—changes each year, it is possible that any growth estimate may be affected by student or program migration. At the district or school level, the impacts can be much larger. The problems we encounter trying to measure improvements in scores are another reason why CDE should consider changing how it releases SBAC data to allow users to arrive at more accurate pictures of growth. They also raise larger questions about how test data are used in the state's accountability programs.

## A Cohort Growth Measure

In this report, we use scale score data to measure growth so that we capture changes in performance for all students. We call this our "cohort" growth measure because we look at the change in scale scores for the same class of students from one year to the next. For instance, our indicator measures the gains 3rd graders in 2015–16 made when they became 4th graders in 2016–17. This yields the most accurate possible measure of growth using the CDE group data.

The SBAC tests' nonlinear design means that growth as measured by scale scores signifies different things in each grade. For instance, the minimum scale score needed for the Met Standard level in mathematics grew 49 points from 3rd to 4th grade, but only 15 points from 6th to 7th (see Figure 1). To create a consistent growth indicator over the grades, we standardize growth scores so they are expressed as a proportion of the amount students are expected to learn during the year. To do this, we report growth as a proportion of the increase in minimum score in the Met Standard level in each grade.

In the final step in calculating our growth indicator, we average the standardized growth over the grades. A growth path of 1 indicates average annual growth in grades 4–8 equals the increase in standards over those grades. Or, to say it another way, students are learning at a rate that would keep them in the Met Standard performance level from one year to the next. (See Appendix A for more detail on how we calculated growth and for data on growth by grade. Technical Appendix B looks at the trends in 3rd and 11th grade scores from 2014–15 to 2016–17. Since SBAC does not test 2nd or 10th graders, there is no way to calculate growth for those grades.)
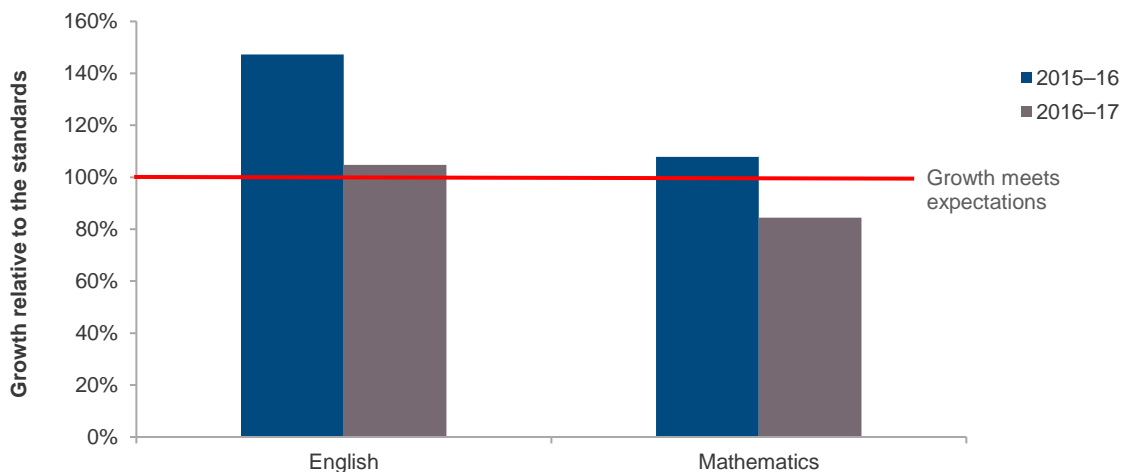
## Growth in Grades 4–8

Figure 8 displays growth estimates for mathematics and English in 2015–16 and 2016–17. The figure shows the amount student achievement grew compared to the increase in the standard for each grade. The red line equals 1, which is the point at which average growth in grades 4–8 matches the increase in standards. The English assessment data show growth of greater than 1, indicating that students learned at rates that exceeded expectations as represented by the standards. English scores in 2015–16 improved 47 percent more than the growth of the standards. But in 2016–17, growth in English was only 3 percent faster than the standards. These data are consistent with the earlier observation that scores are higher on the English test as students move through the grades.

The mathematics scores paint a different picture. Growth in mathematics in 2015–16 was 1.08, or 8 percent higher than growth in the standards. But in 2016–17, growth fell significantly short, registering only 85 percent of the progress students are expected to make each year. As noted earlier, the percentage proficient in mathematics is lower at higher grades. The 2016–17 growth data are consistent with that trend, showing that, on average, students are not learning enough each year to keep pace with the standards.

**FIGURE 8**
Students failed to learn enough mathematics in 2016-17 to keep pace with the standards



SOURCE: California Department of Education and author's calculations.
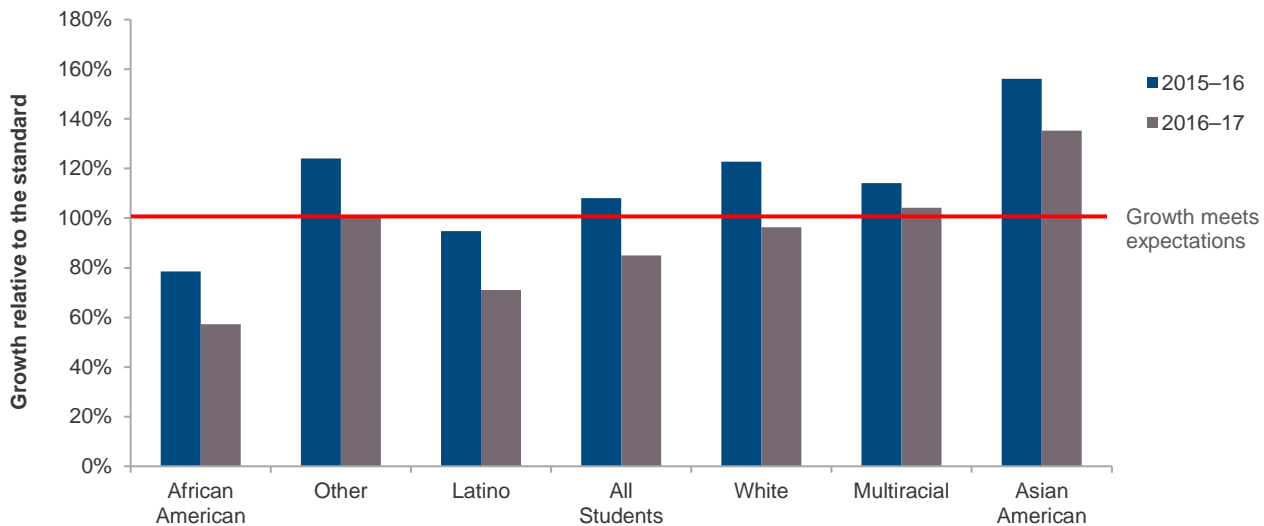
NOTE: Growth is defined as adjacent-grade growth in grades 4–8 (i.e., grade 3 in 2015–16 to grade 4 in 2016–17, grade 4 in 2015–16 to grade 5 in 2016–17) as a proportion of the change in lowest score in "Met Standards" level for that grade.

The much higher growth in 2015–16 compared with 2016–17 represents a very large difference. Growth in English was 42 percent higher and in mathematics 24 percent higher in 2015–16 than in the following year. Does this mean students learned much more in 2015–16 than in 2016–17? We discuss this issue in more detail later in the report.

## Race/Ethnicity

Figure 9 illustrates student achievement growth in five racial or ethnic categories. Again, the red bar marks a growth of 1, indicating that students are growing at the same rate as the standards in grades 4–8. Earlier, we noted that African American and Latino students performed at much lower levels than white and Asian American students. Growth in mathematics reflects a similar trend. The groups are shown in order of the average proportion of students scoring at or above standard in mathematics. African American students had the smallest proportion scoring at proficient levels and also registered the smallest growth. The average African American student gained 78 percent of what was needed to keep pace with standards in 2015–16 and only 57 percent in 2016–17. Latinos fared slightly better, achieving 95 percent and 71 percent, respectively, of the growth in standards in those two years. Among the other three racial or ethnic groups, only Asian students exceeded the amount students are expected to learn in each grade in both years.

**FIGURE 9**
Large differences in annual growth by race and ethnicity in mathematics



SOURCE: California Department of Education and author's calculations.
NOTE: Adjacent-grade growth in grades 4–8 (i.e., grade 3 in 2015–16 to grade 4 in 2016–17, grade 4 in 2015–16 to grade 5 in 2016–17) as a proportion of the change in lowest score in Met Standards score for that grade. Other includes American Indian, Filipino, Hawaiian/Pacific Islander.

These mathematics growth rates are disappointing, illustrating that lower-performing groups are falling further behind each year. As previously noted, African American and Latino students are overrepresented in the low-income student group. These data suggest that LCFF has not yet helped to boost the achievement of the low-income group. Instead, the growth data show that past patterns generally continue, with higher-performing students making larger gains each year than lower-performing students.

Figure 9 also highlights the fact that growth was much larger in 2015–16 for all groups. In addition, the differences in growth in 2015–16 and 2016–17 were virtually the same for each group. The largest difference was for white students. Growth was 27 percentage points lower in 2016–17 for white students than the previous year, falling from 1.23 to 0.96. The smallest differences were for African Americans and Asian Americans, with growth falling for both groups by 21 percent of a year's worth of growth. The consistency of these differences suggests systematic factors may be affecting growth in these two years. We discuss that issue next.

## Putting Growth in Context

The difference between 2015–16 and 2016–17 growth raises the question of how to explain these results. Both English and mathematics gains in 2015–16 far outstripped increases in 2016–17. Growth data for all students and for the main racial and ethnic groups show differences of about 20 percent in the amount students learned in mathematics in 2015–16 compared with 2016–17. The gains were also widespread. When CDE revised accountability standards under LCFF, it noted that more than 80 percent of local education agencies, including districts and some charters, increased their average scores in 2015–16. In 2016–17, only 45 percent posted an increase (CDE 2017b).

These differences raise questions of interpretation. If the significant growth in 2015–16 represents what the system can produce in any given year, then the 2016–17 scores could be considered disappointing. However, several unique factors may have affected the 2015–16 scores. One factor may have been more complete implementation of the Common Core standards, which rearranged when certain material was taught and emphasized "higher order thinking"; that is, using skills and knowledge to solve problems. While the standards were adopted in 2010, implementation started in much of the state only after the previous state tests were discontinued in 2013 (Warren and Murphy 2014). The recession also delayed textbook purchases and investments in teacher training needed to implement the standards.

Students and teachers were also more familiar with the SBAC test in 2015–16. State tests were given on computer for the first time in 2014–15. If first-year scores were depressed because of technical glitches or lack of computer know-how, 2015–16 scores may have increased once those problems were addressed. Similarly, 2014–15 testing helped teachers understand how SBAC measures the Common Core standards. Tests implicitly define what is important in the standards. Knowledge of what is tested on SBAC may have helped teachers align instruction to those priorities, which may have contributed to the 2015–16 jump in scores.

While it would be wonderful to see the K–12 system making large inroads in the achievement gap, it is much more likely that progress will come in small annual increments. The state's previous accountability measure—the Academic Performance Index (API)—identified 5 percent growth as the appropriate annual goal for schools and districts. While API growth is entirely different from our SBAC growth measures, the target reflects the notion that progress comes primarily through moderate sustained increases. Thus, the state needs to take the long view of test scores and create realistic expectations of how fast SBAC scores can improve.

Still, with only two years of growth data, we cannot conclude that the 2016–17 gains are representative of what the system *can* produce. The state is still implementing LCFF. The State Board of Education completed the accountability features of LCFF in 2017 and is still working to complete the support networks that will help educators and administrators improve student results. Higher educational productivity can occur only when teachers and administrators find new ways to help students achieve at higher levels. That kind of change takes time.

## CDE Data Do Not Generate Accurate Estimates of Subgroup Gains

Unfortunately, our analysis shows that using CDE group data to calculate growth estimates for most subgroups of students produces inaccurate measures. We found two major problems with using these data to calculate growth in achievement. First, families move, requiring children to change schools—sometimes in another state. Second, students do not fall into the same program groups each year—EL students are reclassified; family income increases and students are no longer considered low income. The problem is that the group averages that CDE releases do not recognize that student and program migration in these groups prevent the kind of apples-to-apples comparison needed for an unbiased estimate.

Table 1 shows the effect of student and program migration on the growth estimates for the All Students category and the major student subgroups. The table also shows the percentage change in the net size of each student group based on public use data. The All Students statewide growth figure reports the gains using data on all students. (These are the same estimates presented in Figure 8.) The subgroup estimates of statewide growth are calculated as a weighted average for all categories in each subgroup. For example, the Low Income statewide growth estimate is a weighted average of growth for low-income and higher-income students. Similarly, the English Language estimate is the average of all seven language categories. Since all students are included in these weighted averages, mathematically that average should be virtually the same as the All Students category growth estimate. If the All Students growth estimate does not equal the weighted average using subgroup data, changes in the underlying populations must be affecting the accuracy of the subgroup estimate.

As Table 1 shows, the difference between the state-level and subgroup All Students estimates can be positive or negative. For instance, in 2015–16, the size of the low-income subgroup rose 2.3 percent, which causes the growth estimate using the public use data to overstate growth by 4 percent. In 2016–17, the size of that group shrank by 3.2 percent, and the estimates using the public data understate actual growth by 8 percent.

### TABLE 1
Growth measures are unreliable for major subgroups of students

|  | Growth in 2015–16 | | | Growth in 2016–17 | | |
|---|---|---|---|---|---|---|
|  | Change in cohort size | Estimated test score growth | Percent of All Students growth | Change in cohort size | Estimated test score growth | Percent of All Students growth |
| All Students | 0.2% | 1.079 |  | 0.0% | 0.845 |  |
| Race | 0.6% | 1.081 | 100% | 0.8% | 0.841 | 100% |
| Low Income | 2.3% | 1.119 | 104% | -3.2% | 0.775 | 92% |
| English Learners | -18.0% | 0.959 | 89% | -17.8% | 0.720 | 85% |
| Disabled | 6.5% | 1.097 | 102% | 2.9% | 0.851 | 101% |

SOURCES: California Department of Education and author's calculations

NOTES Growth is defined as adjacent-grade growth in grades 4–8 (i.e., grade 3 in 2014–15 to grade 4 in 2015–16, grade 4 in 2014–15 to grade 5 in 2015–16) as a proportion of the change in lowest score in "Met Standards" level for that grade. The change in the cohort size is based on the change in the size of the subgroup of interest (EL, low income, and disabled) for grades 4–8 from one year to the next. Change in the racial cohort size is based on the sum of the absolute value of the change by grade for each racial and ethnic category.

Table 1 shows that changes in subgroup sizes are larger than the All Students data suggest. For the seven subcategories of race and ethnicity, the change remains small and appears to have no impact on growth estimates for those subgroups. For the other three groups in Table 1, changes in the size of the student groups are much larger and the impact of those changes on the accuracy of our growth estimate for low-income and EL students is

significant. Based on our previous discussion of the EL data, it is not surprising that program mobility affects growth data accuracy.

Data for students with disabilities are harder to interpret. Despite fairly large annual changes in the size of the 4–8 grade cohort, the movement of students has only a small effect on the accuracy of the growth estimate. Perhaps few higher-scoring students leave the program. Further analysis of program migration using student-level data is needed to better understand its impact on special education data.

## Movement Can Affect District Data to a Greater Extent than Statewide Data

At the district level, widespread student movement raises additional concerns about using the public use files to estimate growth of the grades 4–8 cohort. The public use SBAC data show that, in districts with more than 1,000 students tested in 2017, the change in the cohort's net number of students from 2015–16 to 2016–17 ranged from about -10 percent to +10 percent. The range is fairly wide for all district sizes. As seen in the statewide data, actual changes in the number of students from 2015–16 to 2016–17 is probably higher than these net numbers suggest.

Because of the significant amount of student mobility in districts, we looked more closely at these growth estimates. To illustrate the impact of student and program migration, we calculated growth scores for one district, San Francisco Unified (SFUSD). The district lost 4.4 percent of its grades 4–8 cohort in 2016–17. Compared to other unified school districts, this amount of change is high, but not unusual, even for large districts.

Table 2 shows changes in the district's student population and estimates of the distortion that subgroup changes created in the district's growth numbers. Overall, SFUSD scores in grades 4 through 8 appeared to grow slightly faster than the increase in standards from one grade to the next. However, this estimate is suspect because the 4.4 percent change in the tested population may signal differences in the types of students in the district compared with 2015–16. The data on race and ethnicity confirm that actual student population changes are larger than the All Students data suggest—and that those underlying changes affect the accuracy of growth calculations.

Changes in the three other subgroups are also large, affecting growth estimates. Notably, estimates based on the weighted average of income, language, and disability groupings show lower district-wide growth than the All Students data. Growth estimates using language and income data are particularly far off, coming in respectively at 18 percent and 20 percent smaller than estimates using the All Students data.

TABLE 2
The changing mix of students affects the accuracy of San Francisco Unified School District growth estimates using group data

| | Growth in 2016–17 | | |
|---|---|---|---|
| | Change in cohort size | Estimated score growth | Percent of All Students growth |
| All Students | -4.4% | 1.026 | |
| Race | 8.7% | 1.08 | 106% |
| Low Income | -18.6% | 0.818 | 80% |
| English Learner | -20.9% | 0.847 | 82% |
| Disabled | 3.6% | 0.950 | 92% |

SOURCES: California Department of Education and author's calculations

NOTES: Growth is defined as adjacent-grade growth in grades 4–8 (i.e., grade 3 in 2015–16 to grade 4 in 2016–17, grade 4 in 2015–16 to grade 5 in 2016–17) as a percentage of the change in lowest score in Met Standards level for that grade. The change in the cohort size is based on the change in the group size for grades four through eight from 2015–16 to 2016–17. Change in the racial cohort size is based on the sum of the absolute value of the change by grade for each racial and ethnic category.

## Similar Problems Affect District Accountability Data

The problems created by student and program migration raise questions about why, in administering California's accountability system, the State Board of Education uses group data for calculating changes in test scores when more accurate measures are available. Importantly, the state refers to the change in scores as "change," not as "growth." Accountability scores in English and mathematics are calculated for students in grades 3–8, not grades 4–8 as in our measure.[8] CDE's change calculation excludes students who have not attended a school or district for most of the school year. These policies mean that students can be included in district accountability scores in one year but not in the other. When the amount of change from one year to the next becomes large enough, comparing scores in the two years does not yield an apples-to-apples comparison.

This problem affects SFUSD's accountability score. We created an All Students estimate of the SFUSD's growth using its dashboard data for mathematics performance by race and ethnicity. The weighted average of the seven categories generated a lower All Students rate, equaling 79 percent of the All Students growth rate calculated using district totals. In addition, the number of students counted in the district's accountability scores were only slightly different than the number of students who took the SBAC test in 2016–17.[9] Therefore, given the impact of student and program migration on the district's accountability score in 2016–17 and the small differences between the tested population and the accountability base, we are unsure exactly what the SFUSD accountability metrics are actually measuring.

We applied the same tests to several other districts and obtained similar results, even when student migration was smaller than in San Francisco. We discussed this problem with CDE staff, who noted that the dashboard calls the difference in scores from one year to the next "change," not growth. They also reminded us that CDE and the State Board are considering revising the way change is calculated.

## California Needs to Reconsider Its Policies for Reporting and Using SBAC Data

Our analysis shows that, for most subgroups, the CDE public use files generally cannot be used to calculate student growth from one grade to the next at the statewide level. At the district level, the impact of changing student attendance and program participation has even more profound effects. The problems we have found with the group data include:

- The performance snapshot of EL students and students with disabilities makes achievement levels look worse than they actually are. Although the data provide an accurate picture of achievement at a moment in time, they fail to recognize the impact of program migration—that is, students enter and leave programs over time, and that this movement affects average scores.

- Student and program migration make group averages almost unusable for assessing student growth from one year to the next. This creates problems for parents, researchers, and policymakers who use the data to understand the progress of students, schools, and districts.

- School and district accountability data are also affected by changes in program participation and student movement. Our analysis raises concerns about the reliability of using changes in English and math scores from the previous year to generate state accountability ratings. Although the State Board partially addresses these issues, the problems appear more widespread than generally recognized.

These findings present two separate, but closely related, issues for the state. The first is how the state's accountability program should address problems with group data. CDE and the State Board are currently

---

[8] Basing district averages on student scores in grades 3–8 ensures some level of distortion because 8th graders in 2015–16 were not tested as they were in 9th grade in 2016–17. Similarly, the 2015–16 average does not contain data on 3rd graders in 2016–17 because they were in 2nd grade that year and not tested. As a result, about a quarter of the students measured in the state's "change" calculation are part of the average for only one year.

[9] Student counts based on race and ethnicity show a difference of 1.9 percent between the tested populations and the accountability populations. Since we have only the subgroup totals, the differences may be larger than these data suggest.

examining options for using student-level data to measure growth—not change—in performance. This is a critical step because it permits the department to develop more accurate indicators of both performance levels and growth. However, the state also needs to revisit other rules, such as when students are included in school and district accountability data.[10] While it has not yet made a commitment to move away from using group data, it is carrying out a thorough review of options in this area. The board plans to decide this matter by October 2018.

The second issue is how the state could present SBAC data to researchers and the public in a way that would overcome the problems that now make the data unusable for measuring achievement growth. CDE's existing public use files provide an easy avenue for obtaining detailed information on SBAC scores, which we applaud. But it is very easy to unintentionally use the data in ways that generate misleading results. Student-level data are available through a CDE application process, but that creates barriers of time and the expertise needed to handle large volumes of complex data. In addition, data requests are subject to CDE approval, which hinges on whether requests are consistent with departmental priorities. In our report, *Increasing the Usefulness of California's Education Data* (Warren and Hough 2013), we suggested ways the state can make data more accessible to the public and school staff. We think this issue needs the department's attention.

# Conclusion

California's test scores in mathematics and English provide important information about the state's K–12 system. Most importantly, the results inform parents, teachers, school administrators, and state policymakers about our children's success in mastering these two basic subjects. In addition, test scores represent the only academic performance measures for students in elementary and middle schools used in the state's K–12 accountability system. Thus, the data are central to evaluating whether schools and districts are performing adequately.

This report uses these publicly available data to explore how students, including the major subgroups of students, have performed during the past three testing cycles. These data create a useful and detailed picture of the current status of achievement. As noted, English proficiency is low in the lower grades and gradually rises through grade 11, when about 60 percent of students test as proficient. By contrast, math test results fall and by grade 11 only about one-third of students score at proficient levels. These trends are consistent with our estimates of the amount students learn each year. For instance, in mathematics, students do not learn enough each year in grades 4 through 8 to keep pace with the standards.

We also examined the performance of student subgroups. Achievement levels for low-income students are much lower for these groups, a finding that has been documented previously. In addition, our regional analysis of low-income student performance showed, on average, very small differences. None of the regions has been successful in boosting the performance of this group. By contrast, the performance of higher-income students varied significantly by region, although the definition of the higher-income group is broad and may not result in comparable groups from one region to the next.

We also found that student score growth was much lower in 2016–17 than in 2015–16. In 2015–16, students made large gains over the previous year in both English and mathematics. However, in 2017, English scores grew only

---

[10] State Board policies continue the inclusion practices put in place due to No Child Left Behind requirements. The new federal law, the Every Student Succeeds Act, appears to provide more leeway to states to determine how to calculate growth.

slightly more than what was needed to maintain proficient performance levels. In mathematics, gains fell far short of keeping pace with standards.

Was 2016–17 simply a disappointing year or was 2015–16 growth unusually large? The answer must await more experience with the SBAC tests. A number of systemic factors—better understanding of the SBAC tests, continued implementation of the standards, and experience with online testing—seems to have boosted 2015–16 scores. It remains to be seen whether the 2016–17 results are representative of what we can expect in the future.

CDE's group data also fall short in important ways. Our ability to use CDE public release files to understand the progress of EL and disabled students is hindered by the movement of students between districts and programs. We showed how EL data understate the success schools are having in helping this group become fluent in English. Moreover, student movement between programs undermines our ability to use subgroup averages to assess year-to-year student growth. This represents a vexing problem for researchers and policymakers because the SBAC tests were designed to measure achievement growth from grade to grade.

School and district data are affected even more significantly than statewide data by changes in program participation and movement of students in and out of districts. These changes make the state's public data on school and district performance unusable for generating district growth estimates. What is more, they also appear to affect the state accountability ratings for English and mathematics performance. The State Board of Education and CDE are exploring whether to change the way accountability measures are calculated, which may address these issues. Moving away from using average group scores would permit the department to develop more accurate indicators of performance levels and growth. However, the state also should revisit other questions, such as attendance rules that determine when students are included in school and district accountability data.

CDE should also reassess how it releases annual SBAC test data. CDE's public use files provide researchers and policymakers a wealth of data on SBAC scores. Because SBAC was designed to measure the annual progress of students, data released by CDE should allow examination of the gains students make each year. Student-level data that would allow researchers to look at growth are available through a CDE application procedure. But that application process creates unnecessary barriers.

Providing educators, policymakers, and the public accurate information about the progress of K–12 students is the central reason why we test students each year. Testing data could provide essential facts to the public about whether LCFF, now in its fifth year of operation, is succeeding in its goal of improving outcomes for low-income, EL, and foster students. Accurate data on the gains made by students with disabilities would inform policymakers on the challenges districts face in educating this group. And parents and community members would have access to better information about the growth of student scores at local schools. Many of these important uses do not require student-level data but can be satisfied with group averages that are adjusted for student movement. To help realize the promise of the SBAC data, CDE should work with researchers and policymakers to revamp its test data release program.

## REFERENCES

California Department of Education. 2017a. *CALPADS Data Guide, A Guide for Program Staff.* Version 9.2.

California Department of Education. 2017b. California School Dashboard, Technical Guide, 2017–18 School Year (November).

Hill, Laura E., Margaret Weston, and Joseph M. Hayes. 2014. *Reclassification of English Learner Students in California.* Public Policy Institute of California.

Hill, Laura and Iwunze Ugo. 2016. *High-Needs Students and California's New Assessments.* Public Policy Institute of California.

McRae, Douglas J. 2017. "Consortium 2017 State-by-State Comparisons," Ed Source.org.

Rodriguez, Olga, Hans Johnson, Marisol Cuellar Mejia, and Bonnie Brooks. 2017. *Reforming Math Pathways at California's Community Colleges*. Public Policy Institute of California.

Ugo, Iwunze and Laura Hill. 2017a. *Student Achievement and Growth on California's K–12 Assessments.* Public Policy Institute of California.

Ugo, Iwunze and Laura Hill, 2017b. *Charter Schools and California's Local Control Funding Formula*. Public Policy Institute of California.

Warren, Paul and Heather Hough. 2013. *Increasing the Usefulness of California's Education Data.* Public Policy Institute of California, August 2013

Warren, Paul and Patrick Murphy. 2014. *Implementing the Common Core State Standards in California*. Public Policy Institute of California.

## ABOUT THE AUTHOR

**Paul Warren** is a research associate at the Public Policy Institute of California, where he focuses on K–12 education finance and accountability. Before he joined PPIC, he worked in the California Legislative Analyst's Office for more than 20 years as a policy analyst and director. He primarily analyzed education policy, but he also addressed welfare and tax issues. Prior to that, he was chief consultant to the state Assembly's committee on education. He also served as deputy director for the California Department of Education, helping to implement testing and accountability programs. He holds a master's degree in public policy from Harvard's Kennedy School of Government.

## ACKNOWLEDGMENTS

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research.