

# Updated Analysis of the Residual Gain Model

**2017–18 Plan**

**Prepared for the California Department of Education by Educational Testing Service**

**Submitted June 11, 2018**

Table of Contents

[Updated Analysis of the Residual Gain Model i](#_Toc516475298)

[Purpose 1](#_Toc516475299)

[Definition of the RG Model 1](#_Toc516475300)

[Analysis 1: Relationships with Student Demographic Variables 3](#_Toc516475301)

[Analysis 2: Relationship with Grade Composition 8](#_Toc516475302)

[Analysis 3: Precision/Reliability 9](#_Toc516475303)

[Analysis 4: Cross-Time Stability of RG 10](#_Toc516475304)

[References 18](#_Toc516475305)

List of Tables

Table 1. Relationships of Each Residual Gain Statistic with Average Prior Scores and Student Covariates—School Level 4

Table 2. Relationships of Each Growth Statistic with Average Prior Scores and Student Covariates—LEA Level 4

Table 3. Correlations Between Each Mean Residual Gain Statistic for ELA and Each Transformed Aggregate Student Demographic Variable—School Level 5

Table 4. Correlations Between Each Mean Residual Gain Statistic for Mathematics and Each Transformed Aggregate Student Demographic Variable—School Level 6

Table 5. Correlations Between Each Mean Residual Gain Statistic for ELA and Each Transformed Aggregate Student Demographic Variable—LEA Level 6

Table 6. Correlations Between Each Mean Residual Gain Statistic for Mathematics and Each Transformed Aggregate Student Demographic Variable—LEA Level 7

Table 7. Average Percentile Rank by School Mean Growth for Each Residual Gain Measure for Schools with Low Percentage (≤ 25%) of Socioeconomically Disadvantaged Students Versus a High Percentage (≥ 75%) 7

Table 8. Correlations to Establish Relationships with Each School Residual Gain Measure and Grade Configurations by Content Area 8

Table 9. Reliabilities of Each Residual Gain Measure by Content Area and Level of Analysis 9

Table 10. Comparing Regression Coefficient Estimates From Each of the Residual Gain Estimates 10

Table 11. Distributions of Residual Gain Measures—School Level 12

Table 12. Distributions of Residual Gain Measures—LEA Level 13

Table 13. Comparing Residual Gain Estimates Over Two Years (2015–16 and 2016–‍17)—School Level 14

Table 14. Cross-classifications by Quintiles\*—School Level 15

Table 15. Comparing Residual Gain Estimates Over Two Years (2015–16 and 2016–‍17)—LEA Level 15

Table 16. Cross-classifications by Quintiles\*—LEA Level 16

Table 17. Comparing Residual Gain Estimates Over Two Years (2015–16 and 2016–‍17)—Grade Level 17

## Purpose

The *California Growth Study[[1]](#footnote-1)*, conducted by Educational Testing Service (ETS) at the request of the California Department of Education, evaluated three candidate growth models for possible inclusion in its accountability system for schools and local educational agencies (LEAs) for the California Assessment of Student Performance and Progress (CAASPP) Smarter Balanced summative assessments in English language arts/‌literacy (ELA) and mathematics. These models included change in distance-to-met, conditional percentile rank of the gain, and residual gain (RG).

In its May 9, 2018, meeting, the California State Board of Education recommended continued investigation of the RG model. This follow-up study represents part of this continued investigation by

1. replicating the three core analyses (relationships with student demographic variables, relationship with grade composition, and precision/reliability) in the *California Growth Study* at the school and LEA levels using the most recent available test score data—residual growth from 2015–2016 to 2016–2017—to determine if similar results hold to those found for residual growth from 2014–‍2015 to 2015–2016, and
2. comparing the aggregated RG estimates for the same matched schools and LEAs across the two sets of time points to evaluate cross-time stability of the growth measure.

This document first briefly defines the model of interest, describes analysis methods, and presents the findings from these analyses.

### Definition of the RG Model

The RG model measures relative or normative student growth. It evaluates students’ current performance relative to students with the same prior performance, providing a measure of how much better or worse a student performed this year than students with the same prior scores performed on average this year. In this study, like in the *California Growth Study*, the RG conditions on prior-year scores in both ELA and mathematics. The underlying regression model used to estimate RGs can accommodate more than a single year of prior scores, but in this context, RG is defined using only a single year.

Specifically, a student’s RG estimate is the difference between the student’s actual current score in ELA or mathematics (ELA2 or Math2) and the student’s expected current score (‍ or ), as shown in equation (1):

  (1)

where the subscript “2” denotes the current year and the subscript “1” denotes the immediately prior year.

For example, in the RG calculation for 2016–2017, year “2” is 2016–2017 and year “1” is 2015–2016. The expected scores are found from regressing the current scores on the prior-year ELA and mathematics scores (ELA1 and Math1):

 (2)

where,

 and  are the intercepts in the ELA and mathematics regressions, respectively; and

 are the regression coefficients for the prior scores in each regression.

Students who receive positive RGs scored higher than would be expected given their prior performance, and vice versa for students who receive negative RGs. Because RGs are found using a regression model, they do not require that the current and prior scores are on the same vertical scale. They do, however, require that all scores are on an interval scale within each grade level. They also require having all or a large representative sample of data available to estimate them.

To aggregate to the school or LEA level, the RGs for individual students linked to a particular school or LEA are averaged. The same business rules for defining schools and LEAs as used for the *California Growth Study* were implemented here, including only reporting results for schools and LEAs with at least 10 students who have an RG estimate.

## Analysis 1: Relationships with Student Demographic Variables

To assess the relationship between RG and student background variables, RG was correlated with average prior performance in ELA and mathematics and with each aggregated student demographic variable of interest. The school and LEA RG estimates were also regressed on (1) average prior performance and the aggregated student demographic variables and (2) only aggregated student demographic variables, to determine the proportion of variance in the aggregated growth statistics, which is explained by the composition of the students in the schools or LEAs. As in the *California Growth Study*, a normalizing transformation was applied to the aggregated student demographic variables before including them in the regression models or estimating the correlations with RG by taking the arcsine square root of the proportion of students in each student group.

Table 1 and Table 2 provide the correlations between RG and the average prior ELA and mathematics performance for schools and LEAs, respectively, as well as the *R*2 values from the two regressions. The 2015–2016 results are those from the *California Growth Study* for growth from academic year 2014–2015 to academic year 2015–2016. The 2016–2017 results are for growth from academic year 2015–2016 (when testing occurred in spring 2016) to academic year 2016–2017 (when testing occurred in spring 2017). They are newly computed and use the most recent test score data available. Table 3 through Table 6 provide the correlations between RG and the aggregated student demographic variables for ELA or mathematics for each of the two time points at the school and LEA levels.

Generally, results are very similar over the two years. While there is some variation in correlations between RG and the aggregated student demographic variables over the two years, it is often for the smaller student groups, such as American Indian or Alaska Native, or foster youth. It is also worth noting that the statistics reported in Table 1 through Table 6 tend to show more variation from year to year at the LEA level than at the school level, even though LEAs generally have larger numbers of students than schools do. The larger variation across years for LEAs is because there are far fewer LEAs than schools, and each of the statistics becomes less stable when computed with fewer units.

Table . Relationships of Each Residual Gain Statistic with Average Prior Scores and Student Covariates—School Level

| Content Area | Year | Correlation with Prior Mean ELA | Correlation with Prior Mean Mathematics | *R*2 for Mean Student Covariates + Mean Prior Scores | *R*2 for Mean Student Covariates Only |
| --- | --- | --- | --- | --- | --- |
| ELA | 2015–16 | 0.13 | 0.11 | 0.08 | 0.08 |
| ELA | 2016–17 | 0.13 | 0.12 | 0.07 | 0.06 |
| Mathematics | 2015–16 | 0.17 | 0.23 | 0.14 | 0.12 |
| Mathematics | 2016–17 | 0.16 | 0.22 | 0.13 | 0.11 |

Table . Relationships of Each Growth Statistic with Average Prior Scores and Student Covariates—LEA Level

| Content Area | Year | Correlation with Prior Mean ELA | Correlation with Prior Mean Mathematics | *R*2 for Mean Student Covariates + Mean Prior Scores | *R*2 for Mean Student Covariates Only |
| --- | --- | --- | --- | --- | --- |
| ELA | 2015–16 | 0.24 | 0.19 | 0.17 | 0.16 |
| ELA | 2016–17 | 0.17 | 0.14 | 0.10 | 0.08 |
| Mathematics | 2015–16 | 0.29 | 0.33 | 0.21 | 0.19 |
| Mathematics | 2016–17 | 0.22 | 0.27 | 0.16 | 0.13 |

As discussed in the *California Growth Study*, even though the correlations in Table 3 through Table 6 are generally weak to modest, they can result in notable differences for schools serving distinctly different populations of students. To illustrate these notable differences, schools are ranked by their RG value and then examined for how different the average percentile rank is between schools serving either a large or small percentage of socioeconomically disadvantaged (SED) students, where “large” is defined as 75 percent or more students classified as SED and “small” as 25 percent or less. Schools with 25 percent or less of their students classified as SED are ranked 12 percentile points higher on average than those with 75 percent or more students classified as SED in ELA and 22 percentile points in mathematics in 2016–2017. As shown in Table 7, these are both similar to the differences found using the 2015–2016 data.

Table . Correlations Between Each Mean Residual Gain Statistic for ELA and Each Transformed Aggregate Student Demographic Variable—School Level

| Student Group | 2015–16 | 2016–17 |
| --- | --- | --- |
| Female | 0.07 | 0.08 |
| American Indian or Alaska Native | -0.12 | -0.08 |
| Asian | 0.08 | 0.09 |
| Native Hawaiian or Other Pacific Islander | -0.05 | -0.04 |
| Filipino | 0.08 | 0.04 |
| Hispanic or Latino | -0.05 | -0.08 |
| Black or African American | -0.09 | -0.09 |
| White | 0.07 | 0.09 |
| Two or more races | 0.05 | 0.07 |
| Missing (Unknown) | -0.02 | 0.03 |
| Ever English learner (EL) | -0.07 | -0.08 |
| Former EL | -0.01 | -0.04 |
| Students with Disabilities (SWD) | -0.05 | -0.12 |
| SED | -0.19 | -0.18 |
| Homeless | -0.05 | -0.09 |
| Foster youth | -0.08 | -0.11 |

Table . Correlations Between Each Mean Residual Gain Statistic for Mathematics and Each Transformed Aggregate Student Demographic Variable—School Level

| Student Group | 2015–16 | 2016–17 |
| --- | --- | --- |
| Female | 0.00 | -0.01 |
| American Indian or Alaska Native | -0.06 | -0.05 |
| Asian | 0.24 | 0.24 |
| Native Hawaiian or Other Pacific Islander | -0.02 | -0.01 |
| Filipino | 0.07 | 0.07 |
| Hispanic or Latino | -0.20 | -0.18 |
| Black or African American | -0.15 | -0.15 |
| White | 0.15 | 0.12 |
| Two or more races | 0.12 | 0.14 |
| Missing (Unknown) | 0.01 | 0.03 |
| Ever EL | -0.10 | -0.07 |
| Former EL | -0.04 | -0.03 |
| SWD | -0.05 | -0.08 |
| SED | -0.28 | -0.25 |
| Homeless | -0.09 | -0.10 |
| Foster youth | -0.11 | -0.13 |

Table . Correlations Between Each Mean Residual Gain Statistic for ELA and Each Transformed Aggregate Student Demographic Variable—LEA Level

| Student Group | 2015–16 | 2016–17 |
| --- | --- | --- |
| Female | 0.08 | 0.05 |
| American Indian or Alaska Native | -0.29 | -0.14 |
| Asian | 0.17 | 0.07 |
| Native Hawaiian or Other Pacific Islander | 0.05 | -0.06 |
| Filipino | 0.10 | -0.01 |
| Hispanic or Latino | 0.02 | -0.03 |
| Black or African American | -0.02 | -0.07 |
| White | 0.00 | 0.04 |
| Two or more races | 0.04 | -0.03 |
| Missing (Unknown) | 0.07 | 0.03 |
| Ever EL | 0.01 | -0.03 |
| Former EL | 0.10 | -0.02 |
| SWD | -0.02 | -0.15 |
| SED | -0.20 | -0.17 |
| Homeless | -0.02 | -0.06 |
| Foster youth | -0.02 | -0.15 |

Table . Correlations Between Each Mean Residual Gain Statistic for Mathematics and Each Transformed Aggregate Student Demographic Variable—LEA Level

| Student Group | 2015–16 | 2016–17 |
| --- | --- | --- |
| Female | 0.01 | -0.05 |
| American Indian or Alaska Native | -0.21 | 0.03 |
| Asian | 0.22 | 0.15 |
| Native Hawaiian or Other Pacific Islander | 0.00 | -0.06 |
| Filipino | 0.04 | -0.03 |
| Hispanic or Latino | -0.20 | -0.19 |
| Black or African American | -0.13 | -0.18 |
| White | 0.18 | 0.13 |
| Two or more races | 0.18 | 0.12 |
| Missing (Unknown) | 0.01 | -0.03 |
| Ever EL | -0.14 | -0.18 |
| Former EL | -0.06 | -0.15 |
| SWD | -0.09 | -0.07 |
| SED | -0.32 | -0.24 |
| Homeless | -0.10 | -0.08 |
| Foster youth | -0.14 | -0.11 |

Table . Average Percentile Rank by School Mean Growth for Each Residual Gain Estimate for Schools with Low Percentage (≤ 25%) of Socioeconomically Disadvantaged Students Versus a High Percentage (≥ 75%)

| Content Area | Year | Percentage SED ≤ 25% | Percentage SED ≥ 75% | Difference\* |
| --- | --- | --- | --- | --- |
| ELA | 2015–16 | 60 | 45 | 14 |
| ELA | 2016–17 | 57 | 45 | 12 |
| Mathematics | 2015–16 | 67 | 43 | 24 |
| Mathematics | 2016–17 | 66 | 44 | 22 |

\*Note: Differences may be inconsistent due to rounding.

## Analysis 2: Relationship with Grade Composition

Given that the growth statistics are aggregated over students in all the grade levels (grades four through eight) that the schools happen to serve, it is useful to determine the extent to which the growth statistic depends on the grade configuration of the schools. While many LEAs serve all of the grade levels, schools have very different compositions, and thus this analysis is focused on only the school level.

The first two rows of Table 8 provide the correlations between school ELA and mathematics RG estimates and the average grade served. The correlations for 2016–17 are similar to those for 2015–16—there is a weak, negative relationship with average grade.

The third row of Table 8 investigates whether the differences between the ELA and mathematics RG estimates for a school are related to school grade configuration. This correlation is near zero for both sets of years.

The fourth row of Table 8 shows that school growth for ELA and mathematics tends to be moderately, positively related, so that schools that tend to have high RG for ELA also have high RG for mathematics. The new results for 2016–17 are very similar to the results from the *California Growth Study* for 2015–‍16.

Table . Correlations to Establish Relationships with Each School Residual Gain Measure and Grade Configurations by Content Area

| Correlation | 2015–16 | 2016–17 |
| --- | --- | --- |
| ELA growth & average grade | -0.11 | -0.09 |
| Mathematics growth & average grade | -0.13 | -0.12 |
| (Mathematics–ELA growth) & average grade | -0.03 | -0.05 |
| ELA growth & mathematics growth | 0.63 | 0.60 |

## Analysis 3: Precision/Reliability

Estimating precision or reliability of aggregated growth is useful, given they rely on test scores with measurement error; but for some schools or LEAs, results may be based on only a modest number of students. Table 9 provides reliability estimates for RG for ELA and mathematics at the school and LEA levels for both sets of years. (See the *California Growth Study* for methodology for estimating these reliabilities.)

The reliability estimates are similar over the two years with the largest difference for ELA at the LEA level. Further investigation revealed this difference is mostly due to some small outlying LEAs that affect the variance estimates used in calculating the reliability.

Table . Reliabilities of Each Residual Gain Measure by Content Area and Level of Analysis

| Year | Content Area | School Level Reliability | LEA Level Reliability |
| --- | --- | --- | --- |
| 2015–16 | ELA | 0.89 | 0.86 |
| 2016–17 | ELA | 0.88 | 0.82 |
| 2016–17 | Mathematics | 0.91 | 0.88 |
| 2016–17 | Mathematics | 0.91 | 0.87 |

## Analysis 4: Cross-Time Stability of RG

The first three analyses reveal that RG performs similarly in terms of relationships with student demographic variables and grade composition as well as precision/‌reliability over the two sets of time points. This last analysis compares RG more directly over the two time points.

First under consideration is the extent to which the underlying regression models used in the calculation of RG are similar over the two time points. Table 10 provides the regression coefficient estimates and *R*2 values for each of the grade-level regressions used to estimate student RG values.

The regression coefficient estimates for prior ELA and mathematics scores in each regression and the *R*2 values are very similar over the two years, showing that the relationships between the current and prior scores are comparable over the two sets of years. In contrast, the intercept estimates differ noticeably, which likely occurs due to volatility in grade-level test score means over these first few years of administration of the CAASSP Smarter Balanced assessments. Instability in the intercepts will likely diminish with future testing as the testing program stabilizes.

Table . Comparing Regression Coefficient Estimates From Each of the Residual Gain Estimates

| Content Area/Grade | Regression Coefficient | 2015–16 Single-Cohort | 2016–17 Single-Cohort |
| --- | --- | --- | --- |
| ELA 4 | Intercept | 46.58 | 81.00 |
| ELA 4 | Prior ELA | 0.66 | 0.65 |
| ELA 4 | Prior Mathematics | 0.34 | 0.33 |
| ELA 4 | *R*2 | 0.72 | 0.72 |
| ELA 5 | Intercept | 116.60 | 64.49 |
| ELA 5 | Prior ELA | 0.66 | 0.67 |
| ELA 5 | Prior Mathematics | 0.32 | 0.32 |
| ELA 5 | *R*2 | 0.74 | 0.74 |
| ELA 6 | Intercept | 244.90 | 244.10 |
| ELA 6 | Prior ELA | 0.65 | 0.65 |
| ELA 6 | Prior Mathematics | 0.26 | 0.26 |
| ELA 6 | *R*2 | 0.72 | 0.74 |
| ELA 7 | Intercept | 189.70 | 215.70 |
| ELA 7 | Prior ELA | 0.63 | 0.63 |
| ELA 7 | Prior Mathematics | 0.31 | 0.30 |
| ELA 7 | *R*2 | 0.74 | 0.75 |
| ELA 8 | Intercept | 302.90 | 286.40 |
| ELA 8 | Prior ELA | 0.65 | 0.64 |
| ELA 8 | Prior Mathematics | 0.24 | 0.25 |
| ELA 8 | *R*2 | 0.75 | 0.76 |
| Mathematics 4 | Intercept | 245.00 | 213.80 |
| Mathematics 4 | Prior ELA | 0.18 | 0.18 |
| Mathematics 4 | Prior Mathematics | 0.74 | 0.75 |
| Mathematics 4 | *R*2 | 0.75 | 0.75 |
| Mathematics 5 | Intercept | 26.90 | 33.53 |
| Mathematics 5 | Prior ELA | 0.21 | 0.19 |
| Mathematics 5 | Prior Mathematics | 0.80 | 0.81 |
| Mathematics 5 | *R*2 | 0.77 | 0.76 |
| Mathematics 6 | Intercept | -145.50 | -139.70 |
| Mathematics 6 | Prior ELA | 0.31 | 0.31 |
| Mathematics 6 | Prior Mathematics | 0.76 | 0.75 |
| Mathematics 6 | *R*2 | 0.76 | 0.76 |
| Mathematics 7 | Intercept | -52.15 | -4.63 |
| Mathematics 7 | Prior ELA | 0.23 | 0.21 |
| Mathematics 7 | Prior Mathematics | 0.80 | 0.80 |
| Mathematics 7 | *R*2 | 0.79 | 0.79 |
| Mathematics 8 | Intercept | -55.77 | -126.90 |
| Mathematics 8 | Prior ELA | 0.28 | 0.25 |
| Mathematics 8 | Prior Mathematics | 0.75 | 0.81 |
| Mathematics 8 | *R*2 | 0.77 | 0.80 |

Next, the extent the distributions of the RG estimates are comparable over the two sets of years is evaluated. Table 11 and Table 12 provide the means, standard deviations (SDs), and selected percentiles for the two sets of years at the school and LEA levels, respectively.[[2]](#footnote-2)

These tables reveal that the overall distributions are similar, particularly at the school level. There is some variation in the distributions at the LEA level that can be attributed to volatility of some small LEAs. However, because there are fewer LEAs than schools, some small LEAs can impact the center and spread of the distributions more on the LEA level than some small schools can similarly affect the school-level distribution.

Table . Distributions of Residual Gain Estimates—School Level

| Content Area | Year | Mean | SD | 5th Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 95th Percentile | Total Number of Groups |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ELA | 2015–16 | 0.44 | 12.43 | -19.70 | -14.72 | -7.06 | 0.70 | 7.98 | 15.46 | 20.30 | 7324 |
| ELA | 2016–17 | 0.69 | 11.78 | -18.49 | -13.67 | -6.47 | 1.14 | 8.09 | 14.66 | 18.97 | 7397 |
| Mathematics | 2015–16 | 0.27 | 12.91 | -20.25 | -15.70 | -7.86 | 0.56 | 8.70 | 16.00 | 20.74 | 7324 |
| Mathematics | 2016–17 | 0.42 | 12.88 | -20.59 | -15.46 | -7.70 | 0.54 | 8.70 | 16.15 | 20.46 | 7397 |

Table . Distributions of Residual Gain Estimates—LEA Level

| Content Area | Year | Mean | SD | 5th Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 95th Percentile | Total Number of Groups |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ELA | 2015–16 | -1.24 | 10.83 | -18.78 | -13.94 | -6.44 | -0.54 | 4.56 | 9.39 | 14.52 | 828 |
| ELA | 2016–17 | 0.14 | 9.32 | -14.20 | -10.10 | -4.97 | 0.04 | 5.41 | 11.27 | 14.71 | 829 |
| Mathematics | 2015–16 | 0.59 | 10.79 | -15.27 | -11.76 | -5.44 | 0.22 | 6.86 | 13.44 | 17.30 | 828 |
| Mathematics | 2016–17 | 1.34 | 10.29 | -13.18 | -9.81 | -4.88 | 0.60 | 7.08 | 13.50 | 19.04 | 829 |

Lastly, RG was compared for the same, matched schools and LEAs over the two time points to determine the cross-time stability of RG, or the extent that the growth statistics for individual schools or LEAs are generally similar across years. Table 13 provides the Spearman rank order correlation between the RG scores for schools over the two years. It shows that for the 7,263 schools that have RG scores over the two years, their scores are correlated 0.35 in ELA and 0.46 in mathematics, which is substantially lower than the correlations often observed between school mean test scores over time. For these data, school mean scores are correlated .96 over the two time points in both subjects.

Table . Comparing Residual Gain Estimates Over Two Years (2015–16 and 2016–‍17)—School Level

| Content Area | Score 1 | Score 2 | Spearman Rank Correlation | Percent Agreement | Total Number of Groups |
| --- | --- | --- | --- | --- | --- |
| ELA | 2015–16 | 2016–17 | 0.35 | 29% | 7263 |
| Mathematics | 2015–16 | 2016–17 | 0.46 | 34% | 7263 |

To help interpret the low, positive correlations in RG over time, the extent to which schools would change classifications if they were classified by their RG value was investigated. Within each year, the schools that appear in both sets of years were rank-ordered by their RG score, and then the percentile rank was obtained by dividing the rank by the total number of matched schools. Schools were then classified into groups by quintiles, e.g., schools with RG scores in the bottom 20 percent were classified in the first quintile, schools with RG scores in the twentieth to fortieth percentile in the second quintile, and so on, with the top, fifth quintile containing schools in the top 20 percent.[[3]](#footnote-3) Finally, each school’s classification over the two years was compared.

Table 14 presents this comparison for ELA in the top half of table and mathematics in the bottom half. It shows, for instance, that only eight percent of schools were ranked in the first quintile in both years, meaning that only eight percent of schools were in the bottom twenty percent by their ELA RG value in both years. If, in contrast, there was high cross-time stability, this percentage would be closer to 20 percent.

The sum of the diagonal of the table yields the percentage of schools classified in the same group over the two years, or the “Percent Agreement” rate reported in Table 13. These agreement rates are also rather low—only 29 percent for ELA and 34 percent for mathematics, meaning that only 29 percent of schools were classified in the same group by their ELA RG score each year. The majority of schools thus change classifications over the two years. In some cases, schools substantially change classifications; two percent of schools, or 150 out of the 7,263, jumped from the bottom twenty percent to the top twenty percent. Table 15 and Table 16 show similar results at the LEA level.

Table . Cross-classifications by Quintiles\*—School Level

| Content Area | Quintiles for 2016 | 1st Quintile 2017 | 2nd Quintile 2017 | 3rd Quintile 2017 | 4th Quintile 2017 | 5th Quintile 2017 |
| --- | --- | --- | --- | --- | --- | --- |
| ELA | 1st Quintile | 8% | 5% | 3% | 3% | 2% |
| ELA | 2nd Quintile | 5% | 5% | 4% | 4% | 3% |
| ELA | 3rd Quintile | 4% | 4% | 5% | 4% | 3% |
| ELA | 4th Quintile | 2% | 3% | 5% | 5% | 5% |
| ELA | 5th Quintile | 2% | 2% | 3% | 5% | 7% |
| Mathematics | 1st Quintile | 9% | 5% | 3% | 2% | 1% |
| Mathematics | 2nd Quintile | 5% | 5% | 5% | 3% | 2% |
| Mathematics | 3rd Quintile | 3% | 5% | 5% | 4% | 3% |
| Mathematics | 4th Quintile | 2% | 3% | 4% | 6% | 5% |
| Mathematics | 5th Quintile | 1% | 2% | 3% | 5% | 9% |

\*Note: Quintiles equal (0, 0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8], (0.8, 1.0] for the percentile rank of the growth score.

Table . Comparing Residual Gain Estimates Over Two Years (2015–16 and 2016–‍17)—LEA Level

| Content Area | Score 1 | Score 2 | Spearman Rank Correlation | Percent Agreement | Total Number of Groups |
| --- | --- | --- | --- | --- | --- |
| ELA | 2015–16 | 2016–17 | 0.34 | 32% | 823 |
| Mathematics | 2015–16 | 2016–17 | 0.44 | 33% | 823 |

Table . Cross-classifications by Quintiles\*—LEA Level

| Content Area | Quintiles for 2016 | 1st Quintile 2017 | 2nd Quintile 2017 | 3rd Quintile 2017 | 4th Quintile 2017 | 5th Quintile 2017 |
| --- | --- | --- | --- | --- | --- | --- |
| ELA | 1st Quintile | 7% | 5% | 2% | 2% | 4% |
| ELA | 2nd Quintile | 6% | 6% | 3% | 3% | 2% |
| ELA | 3rd Quintile | 4% | 5% | 6% | 4% | 2% |
| ELA | 4th Quintile | 2% | 2% | 5% | 6% | 5% |
| ELA | 5th Quintile | 2% | 2% | 4% | 6% | 7% |
| Mathematics | 1st Quintile | 9% | 5% | 3% | 2% | 2% |
| Mathematics | 2nd Quintile | 5% | 6% | 4% | 2% | 2% |
| Mathematics | 3rd Quintile | 3% | 5% | 5% | 5% | 3% |
| Mathematics | 4th Quintile | 1% | 2% | 5% | 5% | 5% |
| Mathematics | 5th Quintile | 2% | 2% | 3% | 5% | 8% |

\*Note: Quintiles equal (0, 0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8], (0.8, 1.0] for the percentile rank of the growth score.

Table 13 through Table 16 indicate that RG exhibits low cross-time stability, but this property is not unique to this particular growth model. Further investigation reveals similar correlations and percentage agreement rates for the two other growth models considered in the *California Growth Study*. Moreover, this property is not unique to California assessments—analysis of a dataset from a large urban school district in a different state found similar magnitudes of cross-time correlations. Low cross-time stability of aggregated growth scores is a known point of concern and can compromise their use in evaluating teachers, schools, or LEAs, or holding them accountable for their students’ progress (e.g., Lash et al., 2016).

The year-to-year instability in the growth measures means that schools with low growth one year might have notably higher growth the next year and vice versa for schools with high growth. Such volatility can make it difficult for educational leaders to use the growth data for driving decisions, as decisions made one year might be contradicted with the next year’s growth data.

It may be more advisable to look for patterns in these scores over several years than to act on their values in a given year. For instance, if educational leaders are interested in taking any action based on aggregated growth scores, they may want to identify schools that remain in the bottom 20 percent over several years as opposed to a single year. Alternatively, to try and improve cross-time stability of the growth scores, a rolling, weighted average could be used for reporting school and LEA growth scores, meaning that school or LEA growth scores from the current year could be averaged with those for the same school or LEA from the prior two (or more) years, perhaps with more weight on the current year growth score. Average growth scores should exhibit more cross-time stability, but empirical analyses would have to be conducted to determine the extent to which they do. Such an approach would require waiting several years into the adoption of the growth model before actually reporting growth scores to schools and LEAs.

Cross-time stability can be improved marginally by disaggregating to the grade level instead of averaging over all grades within a school or LEA. Table 17 provides the correlations and percentage agreement rates at the grade-level. Correlations now range from .33 to .57 instead of .34 to .46 and percentage of exact agreement rates from 31 to 42 percent instead of 29 to 34 percent. Cross-time stability is slightly higher at the school-by-grade or LEA-by-grade levels, but still not at levels that would likely be desirable for high-stakes decisions.

Table . Comparing Residual Gain Estimates Over Two Years (2015–16 and 2016–‍17)—Grade Level

| Content Area/Grade | Level | Spearman Rank Correlation | Percent Agreement | Total Number of Groups |
| --- | --- | --- | --- | --- |
| ELA 4 | School | 0.45 | 33% | 5551 |
| ELA 5 | School | 0.39 | 31% | 5540 |
| ELA 6 | School | 0.52 | 35% | 3903 |
| ELA 7 | School | 0.47 | 35% | 2343 |
| ELA 8 | School | 0.33 | 28% | 2300 |
| Mathematics 4 | School | 0.52 | 35% | 5551 |
| Mathematics 5 | School | 0.47 | 33% | 5540 |
| Mathematics 6 | School | 0.56 | 37% | 3903 |
| Mathematics 7 | School | 0.56 | 38% | 2343 |
| Mathematics 7 | School | 0.51 | 34% | 2300 |
| ELA 4 | LEA | 0.48 | 38% | 730 |
| ELA 5 | LEA | 0.40 | 33% | 736 |
| ELA 6 | LEA | 0.53 | 41% | 722 |
| ELA 7 | LEA | 0.48 | 41% | 684 |
| ELA 8 | LEA | 0.34 | 30% | 680 |
| Mathematics 4 | LEA | 0.54 | 42% | 730 |
| Mathematics 5 | LEA | 0.45 | 37% | 736 |
| Mathematics 6 | LEA | 0.57 | 42% | 722 |
| Mathematics 7 | LEA | 0.57 | 40% | 684 |
| Mathematics 8 | LEA | 0.50 | 37% | 680 |

## References

Lash, A. Makkonen, R., Tran, L., and Huang, M. (2016). *Analysis of the stability of teacher-level growth scores from the student growth percentile model* (REL 2016‑104). Washington, DC: U.S. Department of Education, Institute of Education Statistics, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from <https://ies.ed.gov/ncee/edlabs/>

1. Retrieved from <https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-amard-feb18item01a1-rev.docx> on June 7, 2018. [↑](#footnote-ref-1)
2. Note that the distribution statistics are for all schools and LEAs with RG values available in a year. They are not for the matched set of schools and LEAs—only those that appear in both years—because distribution level statistics are typically presented for the full data used. However, in this case, the majority of the schools and LEAs are common to the two years, so excluding the few new/different schools/LEAs would not substantially change the statistics. [↑](#footnote-ref-2)
3. The choice to use quintiles to classify schools and LEAs is arbitrary. The results are sensitive to the choice of classification criteria, but the general findings would remain the same, that schools and LEAs tend to change categories over time. [↑](#footnote-ref-3)