Validating the SEDA Measures of District Educational Opportunities via a Common Assessment

Megan Kuhfeld,
NWEA

Thurston Domina,
University of North Carolina, Chapel Hill

Paul Hanselman
University of California, Irvine

Abstract

The Stanford Educational Data Archive (SEDA) is the first data set to allow comparisons of district academic achievement and growth from Grades 3 to 8 across the United States, shining a light on the distribution of educational opportunities. This study describes a convergent validity analysis of the SEDA growth estimates in mathematics and English Language Arts (ELA) by comparing the SEDA estimates against estimates derived from NWEA's MAP Growth assessments. We find strong precision-adjusted correlations between growth estimates from SEDA and MAP Growth in math (.90) and ELA (.82). We also find that the discrepancy between the growth estimates in ELA is slightly more pronounced in high socioeconomic districts. Our analyses indicate a high degree of congruence between the SEDA estimates and estimates derived from the vertically scaled MAP Growth assessment. However, small systematic discrepancies imply that the SEDA growth estimates are less likely to generalize to estimates obtained through MAP Growth in some states.

Keywords

Introduction

The Stanford Educational Data Archive (SEDA) district-level achievement and achievement growth data provide unprecedented insights into the spatial and social distribution of educational opportunity across the United States. Recent research using the SEDA data draws attention to the remarkable degree of variation across U.S. public schools in student achievement and achievement growth (Fahle & Reardon, 2018), the magnitude of racial and gender achievement gaps (Reardon, Kalogrides, & Shore, 2018), as well as the effects of the

Great Recession on student achievement (Shores & Steinberg, 2018). SEDA compiles local achievement information from nearly all public school students in the United States and harmonizes these scores into a standardized scale. In doing so, SEDA allows detailed comparisons of school district mathematics and English language arts (ELA) test score *means* and *standard deviations* for students in third through eighth grades across U.S. public school districts. In addition, by comparing these means over time (2008–2009 to 2014–2015 school years), the SEDA data allow for comparisons of third- to eighth-grade *growth* in mathematics and ELA skills across U.S. public school districts.

However, because SEDA infers continuously scaled district achievement scores from the vertically linked National Assessment of Educational Progress (NAEP) scale, these data are based on potentially important assumptions about the cross-state comparability of achievement growth. In this article, we draw on an alternative source of national testing data, namely, the MAP Growth[1] test scores, as a validation check for the SEDA estimates. MAP Growth assessments are administered in a consistent format and with a common scale to approximately 11 million students in over 9,500 schools (NWEA, 2018). As such, they make it possible to estimate achievement levels in MAP Growth participating districts as well as achievement growth estimates to compare with SEDA's estimates. Reardon, Kalogrides, and Ho (2018) used MAP Growth data as a validation check for the district mean scores, but the achievement growth estimates have not been compared previously against an external measure. By investigating the relationship between SEDA and MAP Growth achievement in the subset of U.S. public school districts for which all or nearly all children sit for state-mandated school accountability assessments as well as the MAP Growth tests (approximately 1,000 districts each year), we answer two primary questions: (a) Do the SEDA estimates of achievement growth correspond to those produced using MAP Growth, and (b) to the degree there are discrepancies, which district characteristics explain the lack of agreement?

A Unique Measure of Educational Opportunity Across the United States

By providing detailed measures of achievement levels at multiple points in time for students at several grade levels and across multiple subgroups, the SEDA data facilitate considerations of when and where students acquire key academic skills. SEDA provides measures of average achievement in mathematics and ELA for third through eighth graders in more than 10,000 U.S. public school districts as well as district growth rates from Grades 3 to 8. Reardon (2018) argues that while the third-grade district averages represent educational opportunities available in a community prior to age 9, the growth rates can be thought of as reflecting educational opportunities available to children in middle childhood. SEDA data further allow the estimation of achievement and achievement growth inequalities within district because estimates are also computed separately for each racial/ethnic, gender, free- or reduced-price lunch program participation, special education, and English language learner subgroup for which districts report data.

---

[1]·MAP Growth is an interim assessment for growth from NWEA. It is formally known as Measure of Academic Progress (MAP).

The detailed time and place comparisons that the SEDA data provide make it possible to both identify social disparities and evaluate social and educational policies. SEDA achievement data have been used to assess the degree of variation in achievement across and within U.S. states (Fahle & Reardon, 2018). While student achievement levels do not vary considerably across districts in some states, substantial cross-district variation exists in other states. Fahle and Reardon's (2018) analyses of these achievement level data indicate that cross-district variation is particularly pronounced in states with high degrees of socioeconomic and racial segregation. Elsewhere, Shores and Steinberg (2017) use repeated achievement-level data from SEDA to assess the consequences of the Great Recession on district average achievement. Their analyses indicate that the Recession had a modest but measurable negative effect on student achievement and that this negative effect appears to be particularly pronounced in places in which the recession corresponded to reductions in the resources available to schools.

SEDA's growth measures are particularly valuable because they make it possible to set aside the substantial variability in academic skills that students bring with them at the start of formal schooling (von Hippel, Workman, & Downey 2018) and focus attention on variation in learning opportunities students are exposed to during the late elementary to middle school years. As such, these growth measures provide a plausible indicator of learning opportunities available to youth. Early publications using the SEDA data indicate that these learning opportunities are distributed unevenly across U.S. public school districts. Pooling data across mathematics and ELA, Reardon (2018) suggests that a student in a district that is among the top 5% among all U.S. public school districts experiences the equivalent of approximately 2 years more academic achievement growth between the third and eighth grades than a student in an average-growth public school district. As Reardon notes, these SEDA measures of district-level achievement growth are largely uncorrelated to third-grade achievement levels, suggesting that the cross-district variation of learning opportunities for preschool and early-elementary aged children is largely unrelated to the cross-district variation of learning opportunities for school-aged children.

Key Assumptions in the Production of SEDA Estimates

Given the profound theoretical and practical implications of findings based on the SEDA data, it is essential to ensure that SEDA's estimates of district achievement and growth scores accurately reflect cross-district variation in achievement and educational opportunity. SEDA data are based on district-level reports of the proportion of students who performed at various proficiency levels on achievement tests that states administer to nearly all students in Grades 3 through 8 as part of state and federal accountability policy. As such, SEDA's district mean estimates make assumptions about the location of different cut scores on the NAEP scale across states and years. It is important to understand this scaling process as well as the potential implications of the scaling decisions for the use of the SEDA estimates.

SEDA generates nationally normed estimates of district growth rates for U.S. public school districts using four primary steps:

Step 1: SEDA estimates district means and standard deviations on a standardized within-state scale based on coarsened proficiency count data using a heteroskedastic ordered probit (HETOP) model (Reardon, Shear, Castellano, & Ho, 2017).

Step 2: Next, it uses statewide mean performance on the National Assessment of Educational Progress (NAEP) to map state-specific achievement distributions onto a national distribution.

Step 3: SEDA then transforms district-grade-year-subject scores on the NAEP scale to more readily interpretable grade-equivalent units and standardized scores. On the grade-level scale, the national average fourth-grade NAEP score in 2009 is anchored at 4, and the national average eighth-grade NAEP score in 2013 is anchored at 8. As a result, a one unit different in scores can be interpreted as the national average difference between students one grade level apart (see Reardon, Kalogrides, and Ho, 2018, for a description of other scale transformations that SEDA provides, including the cohort standardized scale that allows for study of absolute changes over time but not absolute comparisons across grades).

Step 4: Finally, SEDA produces district-level within-cohort growth estimates via a hierarchical linear model. Cohort is defined as the set of observations corresponding to sequential grades in sequential years (e.g., a cohort of students who were in third grade in 2009, fourth grade in 2010, fifth grade in 2011, etc.). The model, in which Grades 3 to 8 math and ELA test scores across multiple cohorts are nested within districts, produces an intercept (depending on the centering choices, the intercept represents the district average at Grade 3 or at Grade 5.5, which reflects the first or middle year of the Grade 3–8 range), a within-grade cohort slope, and an across-grade slope estimate for both math and ELA. We describe the hierarchical linear model used to produce the growth estimates in greater detail in the method section.

Each of these steps, which are described in detail by Reardon, Kalogrides, and Ho (2018), rely on a number of assumptions. We discuss these assumptions, as well as the current evidence to support them, in more detail in the following.

Step 1: Assumptions underlying the estimation of state-specific means and standard deviations.—SEDA utilizes student achievement data that is reported in a "coarsened" form, which is to say that student proficiency is reported as counts of students in a district falling in ordered "proficiency" categories (e.g., "basic," "proficient," "advanced"). Reardon et al. (2017) describe how a HETOP model can be used to estimate means and standard deviations of test score distributions from such coarsened data. This approach assumes that the resulting test score distributions are "respectively normal," which is to say that there is a continuous scale for academic achievement for which all within-state district distributions are normal. Using a simulation study as well as real data examples, Reardon et al. demonstrate that the HETOP model produces unbiased estimates of group means and standard deviations, except when group sample sizes are small (e.g., less than 50 students).

Step 2: Assumptions underlying the rescaling of state-specific estimates for national comparison.—Because the coarsened data that SEDA utilizes in Step 1 are based on state-specific achievement tests, SEDA next uses data from nationally representative NAEP study to rescale district means and facilitate cross-state comparison. This rescaling operation assumes that a state's location on the distribution of fourth- and eighth-grade NAEP scores would reflect its location on the distribution of state accountability tests if such a national distribution were observable. To test this assumption, Reardon, Kalogrides, and Ho (2018) compared the NAEP-linked district mean estimates to NAEP Trial Urban District Assessment (TUDA) estimates as well as district mean scores on the MAP Growth assessment. The NAEP TUDA data provide district means and standard deviations on the NAEP scale for 17 large urban districts in 2009 and 20 in 2011 and 2013. Reardon, Kalogrides, and Ho found that students in the TUDA districts scored approximately 0.05 standard deviations higher on the SEDA estimates (using the NAEP-linked scale based on state accountability measures) than the TUDA NAEP assessments, but the average precision-adjusted correlation between the two sets of scores was 0.95. Furthermore, the correlation between SEDA and MAP Growth means was between 0.89 and 0.95. The conclusion from this initial validation was that the linked SEDA estimates can be used to examine variation among districts as well as across grades within districts, but the small amounts of linking error in the estimates do not allow for fine-grained distinctions among districts in different states.

Step 3: Assumptions underlying the construction of the grade-equivalent scale.—The grade-equivalent scale has been widely used when SEDA scores are reported by media outlets (e.g., Badger & Quealy, 2017; Miller & Quealy, 2018) because of its intuitive interpretation. This scale, which is anchored on NAEP national means from 2009 and 2013, relies heavily on the extrapolation of state ranks from years in which no NAEP testing occurred based on trends over NAEP-tested years. This extrapolation hinges on the assumption that NAEP estimates of state-level achievement are stable over time. This assumption is widely held, and analyses regularly use NAEP to study longitudinal trends in achievement (Camera, 2018; Campbell, Hombo, & Mazzeo, 2000). For NAEP to measure trends in achievement accurately, the assessment frameworks must remain sufficiently stable. The NAEP mathematics framework was developed in 1990 and updated in 2005 and again in 2009. The current main NAEP reading scale was developed in 2009, replacing the reading framework that was used for the 1992–2007 reading assessments, though scores are still reported on the same scale. Because the frameworks have been consistent during the period that SEDA provides estimates, it is safe to assume that the NAEP scale is sufficiently stable for SEDA linking.

Step 4: Assumptions underlying the model to produce growth scores.—SEDA growth scores are produced by pooling grade-year-subject average test scores using a hierarchical linear model. The grade-level estimates that feed into this model are based on an important assumption of linearity across grades within a district. Because the NAEP scores are not available for students in third, fifth, sixth, and seventh grades, SEDA must further extrapolate grade-specific state ranks for these grades. SEDA uses a linear interpolation to scale test scores in these grades based on those tested by NAEP (fourth and eighth). The

justification for linking between fourth and eighth grades is that NAEP employs a "cross-grade" scale, which is intended to support score interpretations that span across grades (Thissen, 2012). The cross-grade scale has historically been achieved through cross-grade blocks of items that are administered to both fourth and eighth graders. Though, Thissen points out that the number of cross-grade blocks of items included in the math and reading assessments as well as the overall degree of support for cross-grade interpretations has varied over the past two decades.

The current language on NAEP's website regarding the cross-grade scale in reading provides only moderate support for the use of cross-grade inferences, stating

> Comparisons of overall national performance across grade levels on a cross-grade scale are acceptable; however, other types of comparisons or inferences may not be supported by the available information. Note that while the scale is cross-grade, the skills tested and the material on the test increase in complexity and difficulty at each higher grade level, so different things are measured at the different grades even though a progression is implied. (NAEP, 2018)

That is to say, a score of 250 at fourth grade and eighth grade should not be treated as equivalent because students are tested on different material at each grade. Thissen (2012) summarized the body of evidence around the cross-grade scale as "it is cross grade, but don't push it" (p. 14).

If we are to accept the validity of the NAEP cross-grade scales, the question remains whether it is appropriate to linearly scale test scores across grades. One potential issue with this approach is that a large body of evidence shows that growth decelerates between Grades 3 to 8 on vertically scaled reading and math assessments (Bloom, Hill, Black, & Lipsey, 2008; Dadey & Briggs, 2012). That is to say, the expected growth in third grade is much larger than the expected growth for a student in eighth grade. Because NAEP does not test adjacent grades, it is not possible to quantify whether "one year's growth" on the NAEP scale looks different at various grade levels because only linear estimates of growth from Grade 4 to Grade 8 can be estimated. SEDA's assumption of linear interpolation and its impact on district growth estimates can be evaluated by comparing growth on SEDA-linked estimates to growth on a vertical scale that includes all of the tested grade levels.

Purpose of this Study

The previous validation efforts (e.g., Reardon et al., 2017; Reardon, Kalogrides, & Ho, 2018) provide strong support that the HETOP procedures can be used to produce normally distributed district scale scores for districts with a sufficient number of individuals (Step 1) and that the SEDA district-grade-year-subject mean scores reflect the distribution of district scores observed with TUDA NAEP assessment (Step 2). However, the assumptions needed to justify the growth estimates, particularly the cross-grade linking and linear interpolation (Steps 3 and 4), remain largely untested. To test these assumptions, comparisons with district growth estimates from a vertically linked scale that covers all of the grades reported in SEDA (Grades 3–8) are warranted.

In this study, we investigate whether inferences about district educational opportunities based on the SEDA growth estimates generalize to findings from another national scale of academic growth. First, we assess the convergent validity of the Grade 3 to 8 test scores to evaluate the Reardon, Kalogrides, and Ho (2018) findings using our larger sample of MAP Growth districts. Second, we investigate the convergent validity of the SEDA estimates of Grade 3 to 8 growth, asking: What is the relationship between the across-grade growth estimates from SEDA and those based on the MAP Growth assessments? This analysis is predicated on the assumption that MAP Growth, which is administered on the same Grades 3 to 8 vertical scale using a consistent item bank across the county, provides a useful reference from which to compare estimates from SEDA's constructed scale. Lastly, we explore predictors of the discrepancy between the SEDA and MAP Growth district estimates of Grades 3 to 8 growth.

Data and Methods

SEDA

We use the district Grades 3 to 8 mean scores from SEDA data archive Version 2.1 (Reardon, Ho, et al., 2018). The SEDA district-grade-year-subject estimates are constructed from the federal EDFacts data collection system obtained under a restricted data use license. The EDFacts data include counts of students in each of several ordered proficiency categories (labeled, e.g., as below basic, basic, proficient, and advanced) by school, year, grade, and test subject for all 50 states and the District of Columbia. SEDA district estimates are based on roughly 300 million state accountability test scores on math and ELA tests in Grades 3 through 8 in the years 2009–2015 in every public school district in the United States. For our analyses, we focus on districts that tested between the 2008–2009 school year through 2012–2013.

In this study, we use the district-grade-year-subject (long) estimates in the NAEP metric.[2] We use the mean and standard deviation estimates that are calculated based on all students in the district-grade-year-subject to produce "pooled" (across cohorts) district third-grade achievement and Grades 3 to 8 growth estimates within the districts that had both SEDA and MAP growth scores (more details on the selected districts and the model used to produce growth estimates in the next section).[3] For more details on the SEDA scaling approach, including when and how scores were suppressed prior to reporting, see Fahle et al. (2018).

MAP Growth

For comparison, we use student test scores from NWEA's MAP Growth reading and mathematics assessment. The MAP Growth assessments are computer-based tests typically administered three times a year in the fall, winter, and spring. Each test takes approximately

---

[2] We use the National Assessment of Educational Progress (NAEP) scale rather than grade (within cohort) standardized (gcs) scale to remove any impact from one additional scaling decision (i.e., the choice to standardize the NAEP scores by dividing by the average difference in NAEP scores between students one grade level apart) when comparing the Stanford Educational Data Archive (SEDA) and MAP Growth estimates.

[3] Given that our MAP Growth data did not fully overlap with the SEDA districts in terms of national coverage and years of data collected, we chose to produce custom SEDA intercept and growth scores rather than use the pooled estimates from SEDA v2.1 ("SEDA_geodist_poolsub_GCS_v21.csv"). However, we provide a comparison of our custom SEDA estimates and the SEDA v2.1 estimates in the online supplemental material.

40 to 60 minutes depending on the grade and subject area. Students respond to assessment items in order (without the ability to return to previous items), and a test event is finished when a student completes all the test items (typically 40 items for reading and 50 for math). Test scores, called RITs, are reported in an item response theory (IRT)-based metric.

MAP Growth is used by more than 11 million students at over 9,500 schools and districts within the United States and internationally. These assessments are specifically designed for measuring year-to-year academic growth on a consistent scale across grades and settings. Growth measures from these data require comparatively fewer assumptions than SEDA. While all measures of growth depend on the assumption of an equal-interval scale,[4] MAP Growth requires none of the additional assumptions related to deriving continuous estimates from coarsened data, state-equating, or grade-level interpolations. Additionally, unlike most vertically scaled assessments that use a small set of linking items to establish across-grade links, MAP Growth administers items from cross-grade item pools so the assessment can adapt to each student's instructional level, even if that is outside the content standards tied to the student's grade. Thus, MAP Growth is a useful reference for examining the convergent validity of the SEDA estimates.

Analytic Sample

We construct an analytic data set of districts that tested both on the state accountability assessments used by SEDA and MAP Growth. Based on the SEDA data, all observations reflect school district-years for public districts with test scores between 2008–2009 and 2012–2013. We adopt SEDA's definition of district membership, which includes all traditional public and charter school students who attend schools inside a district's geographic borders. This definition implies the inclusion of charter schools in the estimation of district achievement and growth scores. While this definition is appropriate for SEDA, because charter schools must administer the state accountability tests on which SEDA scores are based, charters typically do not participate in district decisions to assess students via the MAP Growth, and therefore MAP Growth estimates may exclude some charters. To ensure comparability between the student populations represented in the independent assessments, we limit the sample to districts with a consistent set of students and schools reporting SEDA and MAP Growth scores. We exclude any district-grade-subject-year in which the ratio of MAP to SEDA students is below 0.9 or above 1.1. The resulting data set includes 1,895 unique districts, representing about 18% of the public school districts in the United States with SEDA scores. Table 1 reports sample size in each year and grade; the analytic sample represents 6% to 18% of districts in SEDA within a given grade and year. Additional sample restrictions for the estimates of educational opportunity are described in the "Educational Opportunities Scores" section.

[4.] The topic of whether an achievement scale can truly have an equal-interval vertical scale has been discussed by many researchers (see Briggs, 2013, for a recent review). Equal-interval properties mean that the same difference of X units (e.g., ability) will produce the same difference in observed behavior (e.g., the probability of getting an item correct) at all points of the scale. A vertical scale most often refers to a common scale that is built by linking multiple tests that measure a similar construct spanning a developmental continuum. The equal-interval properties of Rasch-based measure of ability can be tested using a framework known as the theory of conjoint measurement (see the recent work, e.g., by Briggs, 2013; Domingue, 2014; Karabatsos, 2001). Using this framework, Thum (2018) recently tested whether the MAP Growth assessment displayed equal-interval properties. He examined item responses of third through fifth graders from a mid-Atlantic state using a generalized linear model and found a pattern of score magnitudes that is consistent with additive conjoint assumptions of an equal-interval scale within and between the grade levels.

District Characteristics

We utilize the rich set of district and community characteristics reported by SEDA for the second set of analyses. The district-level covariates were calculated by SEDA using data from the Common Core of Data (CCD) at the National Center of Education Statistics, the School District Demographics System (SDDS), and the U.S. Census Bureau's American Community Survey (ACS). We focus on measures of (a) district demographics: percentage of White, Black, and Hispanic students, a composite socioeconomic status (SES) variable, and percentage of English language learners (ELL); (b) district inequality: White-Black segregation, Gini coefficient, and the 90/10 income ratio; and (c) organizational resources: total enrollment, pupil-teacher ratio, and revenue per pupil. The calculation of the covariate variables is described in the SEDA technical documentation (Fahle et al., 2018).

Table 2 provides a summary of district characteristics in the full SEDA data and the set of districts used in these analyses. On average, the districts used in our analyses serve fewer children on average and higher average SES families across the years of the study than the full set of districts with SEDA scores.

Educational Opportunity Scores

Following the procedures outlined by Reardon, Kalogrides, and Ho (2018), we calculate educational opportunity growth scores using separate models for the SEDA and MAP Growth estimates.[5] To produce pooled average score and across-grade slopes, we fit a hierarchical linear model (HLM) with scores from each grade/year/subject nested within each district. Let y_{dygs} be either SEDA or MAP Growth mean score for students in district d in year y, grade g, and subject s. Separately for each score estimate y_{dygs}, we fit the following model

$$y_{dygs} = \begin{bmatrix} \beta_{0md} + \beta_{1md}(\text{cohort}_{dygs} - 2004.5) \\ + \beta_{2md}(\text{grade}_{dygs} - 3) \end{bmatrix} M_s + \begin{bmatrix} \beta_{0ed} + \beta_{1ed}(\text{cohort}_{dygs} - 2004.5) \\ + \beta_{2ed}(\text{grade}_{dygs} - 3) \end{bmatrix} E_s + u_{dygs}$$

$$\beta_{0md} = \gamma_{m00} + \upsilon_{0md}$$
$$\beta_{1md} = \gamma_{m10} + \upsilon_{1md}$$
$$\beta_{2md} = \gamma_{m20} + \upsilon_{2md}$$
$$\beta_{0ed} = \gamma_{e00} + \upsilon_{0ed}$$
$$\beta_{1ed} = \gamma_{e10} + \upsilon_{1ed}$$
$$\beta_{2ed} = \gamma_{e20} + \upsilon_{2ed}$$

$$u_{dygs} \sim N(0, \sigma^2); \begin{bmatrix} \upsilon_{0md} \\ \vdots \\ \upsilon_{2ed} \end{bmatrix} \sim MVN(0, \tau^2),$$

where M_s and E_s are indicators of subject ($M_s = 1$ if the subject is math, and $E_s = 1$ if the subject is reading; 0 otherwise). As in SEDA's model, we define a cohort as the spring of the year in which a group of students would have been in kindergarten (so that cohort$_{dygs}$ = year – grade). The β_{0sd} parameter represents the mean test score in subject s (where m = math, e = reading) and district d for Grade 3 and Cohort 2004.5 (the centered grade and cohort

[5.]It is important to note that SEDA uses a precision-weighted random-coefficient models, where the uncertainty in y_{dygs} is brought into the model in the form of known variance of a residual term, while our approach does not include separate residual terms for model error and measurement error.

values). The β_{1sd} parameter represents the average within-grade (cohort-to-cohort) change per year, and the β_{2sd} parameter represents the within-cohort change per grade in average test scores for subject s and district d. In this model, the variance term σ^2 and matrix τ^2 are estimated.

The Reardon, Kalogrides, and Ho (2018) model specification described previously uses a linear functional form to describe growth across grade levels, which is a logical selection for SEDA given the linear cross-grade interpolation used to scale the scores. However, MAP Growth, like many other vertically scaled assessments (Dadey & Briggs, 2012), has been demonstrated to show score deceleration (e.g., grade-to-grade growth that decreases across grade levels; Thum & Hauser, 2015). Given that a linear functional form may not accurately capture the grade-to-grade growth patterns with MAP Growth, we also fit an HLM using MAP Growth scores that contained a quadratic fixed effect term for both math and reading.

The parameter estimates from the SEDA and MAP Growth (linear and quadratic) HLMs are shown in Table 3. Because the SEDA and MAP Growth assessments are on a different scale, the parameter estimates are not directly comparable. However, we can see the magnitude of variability in the growth estimates (measured by the standard deviation) relative to the district growth mean estimate is similar between SEDA and MAP Growth. This finding is consistent with Reardon (2018), who demonstrated that there is large degree of between-district variation in growth rates. Both quadratic terms in the nonlinear MAP Growth HLM were negative and statistically significant, confirming that growth is decelerating across grade levels. Given that the quadratic model best captures the patterns of grade-to-grade growth observed with the MAP Growth assessment, we present the results from this model for the subsequent analyses comparing SEDA and MAP Growth.

Table 3 also displays the maximum likelihood estimates of the correlation among the third-grade intercept and average Grades 3 through 8 growth for math and ELA. The math and ELA average third-grade scores are highly correlated in both SEDA and MAP Growth, but the correlation between the math and ELA district growth scores is higher for MAP Growth (.85) than SEDA (.73). We also note that the reliability of the growth estimates from these models is poor for both SEDA and MAP Growth, ranging from .49 to .65. This reliability is lower than the reliability of growth estimates reported in Table 2 of Reardon (2018), which were based on much larger set of districts and a slightly wider span of years.

We derive empirical Bayes (EB) estimates of third-grade achievement and Grades 3 to 8 linear growth from the SEDA linear model and the MAP Growth quadratic model. The EB estimates are the sum of the fixed effects estimate and the EB shrunken estimate of the random effect (e.g., $\hat{\beta}_{kd} = \gamma_{k0} + v_{kd}^{EB}$ for the kth parameter). The EB estimates are biased toward the fixed effect estimate as a function of the unreliability of the residual. We estimate uncertainty in the EB estimate as the square root of the posterior variance of the EB estimates. Following SEDA's procedure, we do not analyze the district estimates of the intercept or growth parameters if the reliability is below 0.7.[6] The reliability of estimate $\hat{\beta}_{kd}$ is computed as $\frac{\hat{\tau}_k{}^2}{\hat{\tau}_k{}^2 + \hat{V}_{kd}}$, where $\hat{\tau}_k{}^2$ is the *kth* diagonal element of the estimated τ^2 matrix

(the estimated true variance of β_{kd}) and $\hat{V}_{kd}$ is the square of the estimated standard error of $\hat{\beta}_{kd}$. As a result of the reliability restrictions, 1,850 districts out of the analytical sample of 1,895 districts were used for the comparisons of the third-grade achievement scores, whereas 1,297 districts in math and 1,019 districts in ELA were used in the comparison of the MAP and SEDA Grades 3 through 8 growth estimates (see Appendix Table A1 for a comparison of the analytical sample and the districts that were dropped).

Results

Comparing District-Grade-Subject-Year Mean Scores

Before calculating growth estimates, we first compare the estimated SEDA grade-year-subject district means with the MAP Growth district means. We calculate precision-adjusted correlations between the SEDA estimates on the NAEP scale and the MAP Growth district means. These correlations within the 2011–2012 school year are shown in Figure 1 alongside scatterplots showing the relationship between the two sets of estimates by grade and subject. Patterns of correlations did not vary greatly across school years. To avoid confusion regarding the different scales used by SEDA and MAP Growth, district means are standardized within year-subject prior to plotting. Additionally, the distribution of scores for each grade separately are shown on the axes of the figures.

The correlations between the SEDA and MAP Growth grade-level estimates reported in Figure 1 range from .85 to .92. Our reported correlations are slightly lower than the SEDA/MAP Growth correlations reported by Reardon, Kalogrides, and Ho (2018), which ranged from .90 to .95, but nonetheless show high convergence between the two sets of estimates. Figure 1 also allows for a comparison of the distribution of district mean scores across grades. SEDA's score distributions (shown on the y-axis) are mostly uniform by grade with almost equivalent mean gaps across grades. By contrast, we observe relatively wide spacing in MAP Growth scores (shown on the x-axis) between the third, fourth, and fifth grades, while the distributions of the later grades are generally wider and more overlapping. We attribute this pattern to SEDA's linear interpolation of scores between fourth and eighth grade, which ensures the district means in the grades in between are equally spaced. MAP Growth, on the other hand, has vertical scale across Grades 3 to 8, so the grade-to-grade differences on the RIT scale are directly observed rather than interpolated.

Another way of looking at the differences in score distribution between SEDA and MAP Growth is to plot score trajectories across time for specific cohorts; tracking district mean scores for third graders in one spring, fourth graders in the next spring, and so on. Figure 2 shows the estimated third through seventh–grade score trajectories in math for cohorts in a randomly selected set of 25 districts. Given the years of data included in the study (2008–2009 to 2012–2013), this represents the longest trajectory directly observed within a cohort. The panel on the left tracks test score trajectories in MAP Growth, and the panel on the right

---

[6.]We chose to use the .70 reliability threshold to remove noisy empirical Bayes (EB) estimates following SEDA's official suppression rules. However, we also reran our analyses applying a stricter reliability cutoff (.80) and found that the correspondence between the SEDA and MAP Growth estimates only marginally improved.

tracks SEDA score trajectories for students in the same 25 districts. The average trajectory for each scale is also shown as a solid black line. Not surprisingly, the MAP Growth scale shows that growth between grades begins leveling off by the end of elementary school, whereas the average trajectory of SEDA scores is linear (due to the linear interpolation used by SEDA to project scores to grades not tested by NAEP (third, fifth, sixth, and seventh grades). Furthermore, within-district variation in the year-to-year gains on the SEDA scale appears somewhat more variable than seen in the MAP Growth scores. On the SEDA scale, some districts show gains followed by losses or temporary flattening, while those same districts show moderate but consistent gains on MAP Growth. It is unclear if the SEDA within-district variability in year-to-year growth reflects the true variability in growth on the state assessments or is (at least in part) an artifact of the linear cross-grade scale interpolation.

Comparing the Educational Opportunity Estimates

How do MAP and SEDA measures of learning opportunities compare for early (mean third-grade scores) and middle (Grades 3–8 linear growth) childhood? Figure 3 displays the relationship between the SEDA and MAP Growth estimates of Grade 3 achievement in Panel a and SEDA and MAP Growth estimates of achievement growth between third and eighth grades in Panel b. District estimates in these bubble plots are weighted by their reliability and color-coded by SES, and a 45° line is overlaid on each plot. All district estimates with reliability less than 0.70 are excluded from these analyses. The precision-weighted correlation between the SEDA and MAP Growth estimates of the district Grade 3 scores is 0.98 in math and 0.97 in ELA. As reported by Reardon (2018), the district Grade 3 estimates shown in Panel a of Figure 3 are associated with SES, with high SES districts mostly clustered at the upper end of the Grade 3 score distribution for both SEDA and MAP Growth. The two plots in Panel a of Figure 3 document a high degree of alignment between SEDA and MAP Growth in third-grade achievement.

However, as the bubble plots reported in Panel b of Figure 3 make clear, SEDA estimates of achievement growth during the elementary and middle school years do not align quite as neatly with MAP Growth linear estimates of achievement growth (from the quadratic model) during the same period. The precision-weighted correlations between the Grades 3 through 8 growth EB estimates from the separate SEDA and MAP Growth models are .90 in math and .82 in ELA. Panel b of Figure 3 clearly shows that while there is a strong association between the two sets of growth estimates, there are some discrepancies between the estimates on the two scales. For example, approximately 3% of districts are within one standard deviation of the mean on one scale but more than two standard deviations above or below the mean on the other scale.

Examining the Discrepancies in the Grade 3 Through 8 Growth Estimates

The analyses reported in Table 4 explore the divergence between SEDA and MAP Growth estimates of elementary and middle school achievement growth. Correlations between the SEDA and MAP Growth EB estimates of third-grade scores and Grades 3 through 8 growth and a set of district characteristics are presented. Not surprisingly, given the tight alignment between third-grade SEDA and MAP Growth scores, the pattern of correlations between the

Grade 3 scores and the district covariates are very consistent across the SEDA and MAP Growth estimates. In both math and ELA, the SEDA and MAP Growth estimates of the Grade 3 scores are strongly positively associated with SES and the percentage of White students in the district and negatively correlated with the extent of racial and income segregation in the district. However, the SEDA and MAP Growth Grades 3 through 8 growth estimates diverge somewhat in their relationships with the district covariates in ELA. For example, the SEDA Grades 3 through 8 ELA growth estimates are not significantly correlated with SES, while the MAP Growth estimates show a small positive association with SES.

The correlations reported in Columns 5 and 10 of Table 4 investigate the degree to which differences between SEDA and MAP Growth third to eighth–grade achievement growth estimates are related to district characteristics. For each district, we obtain an estimate of discrepancy between the SEDA and MAP Growth between Grades 3 and 8 growth. Because the growth estimates from SEDA and MAP Growth are not on the same scale, we first standardize each set of estimates before calculating the difference $\hat{\beta}_{MAP,d} - \hat{\beta}_{SEDA,d}$. Scores with a positive discrepancy value are systematically higher on MAP than SEDA, and scores with a negative value fall lower in the distribution on MAP Growth than SEDA. Table 4 presents the correlations between growth discrepancy scores and the district characteristics. In math, the discrepancy scores are only very weakly correlated with district characteristics such as percentage of English language learners (.06) and the percentage of White students in the district students (–.06), with districts containing a higher percentage of White students more likely to show a negative discrepancy (SEDA scores greater than MAP Growth). In ELA, the discrepancy scores are positively correlated (.15) with SES, implying higher SES districts tend to have higher scores on MAP Growth than SEDA, raising the possibility that SEDA may provide overly optimistic estimates of ELA achievement growth in socioeconomically disadvantaged districts than what would be obtained on other assessments.

Another discrepancy question we asked is whether there are state-by-state differences in the alignment between SEDA and MAP Growth scores. Since the HETOP procedure is conducted state by state and year by year, it is possible that the cumulative linking error results in noisier estimates in some states than others. We estimate correlations between the SEDA and MAP Growth Grade 3 and growth estimates within the 17 states with at least 20 districts represented. Table 5 shows the relationship between the estimated MAP Growth-SEDA correlations in math and ELA. With respect to third-grade mean scores, the correlations between assessments are high overall (range, .53–.95; average = .84). The ordering among states is also stable across subjects. For example, the correlations are lowest in Kansas and highest in Illinois for both subjects, suggesting general features of these assessment and policy contexts account for both discrepancies.

For growth-based measures of middle childhood opportunities, the correlations between SEDA and MAP Growth estimates are lower and more variable. Correspondence tends to be higher in mathematics (correlations range, .50–.82 across states; average = .69) than ELA (correlations range .25–.77; average = .57). Finally, it is notable that state-level consistency is almost always comparable to or lower for growth than third-grade means. For instance,

correlations are similarly high in Minnesota for both measures and subjects, while the growth measure is less aligned with the MAP Growth benchmark in states including Michigan, Indiana, and New Jersey. Thus, the SEDA growth measures seem less likely to generalize to trends obtained with MAP Growth in some settings.

Discussion

This article examines the convergent validity of the SEDA district estimates of student achievement and growth. SEDA achievement and growth scores have already received national news coverage for their ability to highlight geographic variation in educational opportunity as well as gender and racial/ethnic achievement gaps (Badger & Quealy, 2017; Miller & Quealy, 2018). We use aggregated district data from MAP Growth, a nationally administered interim assessment of math and ELA, to provide evidence of convergent validity for the SEDA Grades 3 through 8 growth estimates as measures of educational opportunity.

We find very strong convergence between SEDA and MAP Growth district estimates within each grade/year and strong convergence between the Grades 3 through 8 growth estimates. Consistent with Reardon, Kalogrides, and Ho (2018), we find strong correlations between the district achievement scores across grades and years. This finding adds to the prior validation efforts that suggest SEDA's NAEP-linked district test score results may be useful in analyzing national variation in district-level academic performance, suggesting that SEDA's technique for interpolating scores across years and grades in which NAEP data are not available is convergent with another assessment's results.

We note, however, that our analyses of grade-to-grade achievement growth patterns suggest that SEDA data make different assumptions about the shape of students' average growth trajectories than most vertically scaled assessments. Because SEDA uses a linear interpolation to scale student scores across grades for which NAEP data are not available, SEDA estimates suggest that on average, students experience similarly sized achievement gains in each academic year between third and eighth grades. Test score data from MAP Growth and other vertically scaled achievement tests suggest that this is not the case (Bloom et al., 2008; Dadey & Briggs, 2012). Instead, these data indicate that student achievement growth rates slow, on average, in the upper elementary and middle school years.

Our findings demonstrate that SEDA's estimates of Grades 3 to 8 growth, which are dependent on linear interpolations of NAEP's cross-grade vertical scale, mostly generalize to cross-grade growth patterns observed with other student assessments. The precision-weighted correlations for the growth estimates are strong (.90 in math and .82 in ELA). Correlations in this range indicate that while the overall pattern across districts is consistent, there are a small proportion of districts that appear to show average growth on MAP Growth but are several standard deviations above average on SEDA, and vice versa. Additionally, we find that correspondence between the SEDA and MAP Growth estimates of Grades 3 to 8 growth is lower in some types of districts and some states. In ELA, the discrepancy between SEDA and MAP growth is weakly correlated with SES and percentage of ELL students in

the district. Additionally, the correspondence between the growth estimates is lower in some states than others, implying the divergence in the growth estimates is not random.

There are multiple possible explanations for the discrepancies in the Grades 3 through 8 growth. First, while SEDA is based on federally mandated state accountability test scores, MAP Growth is an optional district-chosen assessment that is typically administered for lower-stakes purposes. Although our analytic sample only includes the approximately 10% of U.S. public school districts in which virtually all students contribute to SEDA and MAP Growth achievement estimates, it is possible that both the population of students tested and timing of the tests in the spring differ between SEDA and MAP Growth. Second, while the content and difficulty of the state accountability tests vary across states, the MAP Growth assessment is based on the same blueprint and overlapping item pool across states. While there is strong overall alignment between SEDA and MAP Growth district-grade-year-subject scores, our exploratory analysis of state-specific difference in third-grade scores revealed some states have lower alignment than others, which could imply that discrepancies are related to local conditions.

There are limitations to this study. The districts administering MAP Growth are not a representative set of districts within SEDA. On average, the districts included in this study have higher SES and lower total enrollment than the national sample of districts. The degree that the divergence in growth scores is explained by SES is possibly underestimated by this study due to the lack of very low SES districts. Additionally, because we used fewer years of data and far fewer districts than SEDA used, our EB estimates of growth are generally noisier than those reported by SEDA. We limited the subsequent analyses that compared the district EB estimates to only those with sufficient reliability (>.70) to avoid making conclusions off of noisy estimates but at a cost of losing a sizable portion of the districts (see Table A1 in the Appendix).

We chose to use MAP Growth to study the convergent validity of the SEDA estimates because it represents the most widely used cross-state measure of academic achievement during the school years (e.g., 2008–2009 to 2014–2015) in which district data are available from SEDA. While for this study we used district-aggregated estimates to mirror the SEDA data preparation steps, future research could use student-level MAP Growth data to examine the sensitivity of growth estimates to the assumption of cohort stability. By tracking individual students, we can test the degree to which a cohort's mean score improves in part due to low-achieving students leaving (or being forced out of) a district. Additionally, to build a stronger body of convergent validity evidence, future research should investigate whether growth estimates produced from state longitudinal systems show similar convergence with the SEDA growth estimates.

Nonetheless, we believe our findings have important implications for scholars interested in using the SEDA district-level estimates to understand the distribution of learning opportunities across U.S. public school districts. Based on our findings, we believe that scholars should have a great deal of confidence in the validity of the SEDA estimates of third-grade achievement levels in U.S. public school districts. Further, although our findings suggest that the SEDA scores distinguish between high- and low-achieving districts in

subsequent grades akin to another assessment, we find that these data may lead to different conclusions about typical patterns of grade-to-grade achievement growth than those drawn from vertically scaled assessments. This limitation is likely not consequential for most SEDA users, though it may limit the data's potential for speaking to issues around middle school transition or other grade-specific achievement trajectories. Finally, although our findings indicate that while the SEDA's growth estimates are capturing similar patterns of Grades 3 to 8 growth as MAP Growth, a small percentage of low-growth districts in SEDA would be identified as high-growth based on the MAP Growth assessment, and vice versa. These discrepancies are related to observable district characteristics, particularly in ELA, where compared to MAP Growth, the SEDA data appear to modestly underestimate the degree of achievement growth in high-SES districts and overestimate the degree of achievement growth in low-SES districts. While these discrepancies do not appear to be large enough to compromise inferences from most analyses based on SEDA data, they should raise cautions about the generalizability of the SEDA growth estimates to other assessments for inferences about learning opportunities in middle childhood.

Acknowledgmentss

Appendix

TABLE A1

Comparison of the Average District Characteristics for Districts Who Were Dropped (Due to Low Reliability of the EB Estimates) vs. the Districts Used in the SEDA-MAP Growth Comparison of Achievement and Growth

| | Grade 3 Estimates | | | | | | | | Growth Estimates | | | | | | | |
| | ELA | | Math | | | ELA | | Math | | | ELA | | Math | | |
	Dropped	Retained	Dropped	Retained	Dropped	Retained	Dropped	Retained	Dropped	Retained	Dropped	Retained	Dropped	Retained
Percent White	0.83	0.80	0.83	0.80	0.80	0.80	0.79	0.81
Percent Hispanic	0.10	0.09	0.07	0.09	0.09	0.09	0.09	0.09
Percent Black	0.03	0.06	0.06	0.06	0.07	0.06	0.07	0.06
SES	0.22	0.30	0.25	0.30	0.19	0.39	0.19	0.35
Percent ELL	0.03	0.04	0.03	0.04	0.03	0.04	0.04	0.04
White-Black Segregation	0.04	0.03	0.04	0.03	0.03	0.03	0.03	0.03
Gini coefficient	0.36	0.35	0.35	0.35	0.35	0.34	0.35	0.34
90/10 income ratio	8.35	7.59	7.77	7.59	7.91	7.32	7.87	7.47
Total enrollment	2,111	1,293	1,346	1,298	1,289	1,307	1,332	1,284
Pupil-teacher ratio	16.11	15.01	14.63	15.02	15.10	14.94	15.13	14.96
N	15	1,880	47	1,848	876	1,019	598	1,297

(a)

(b)
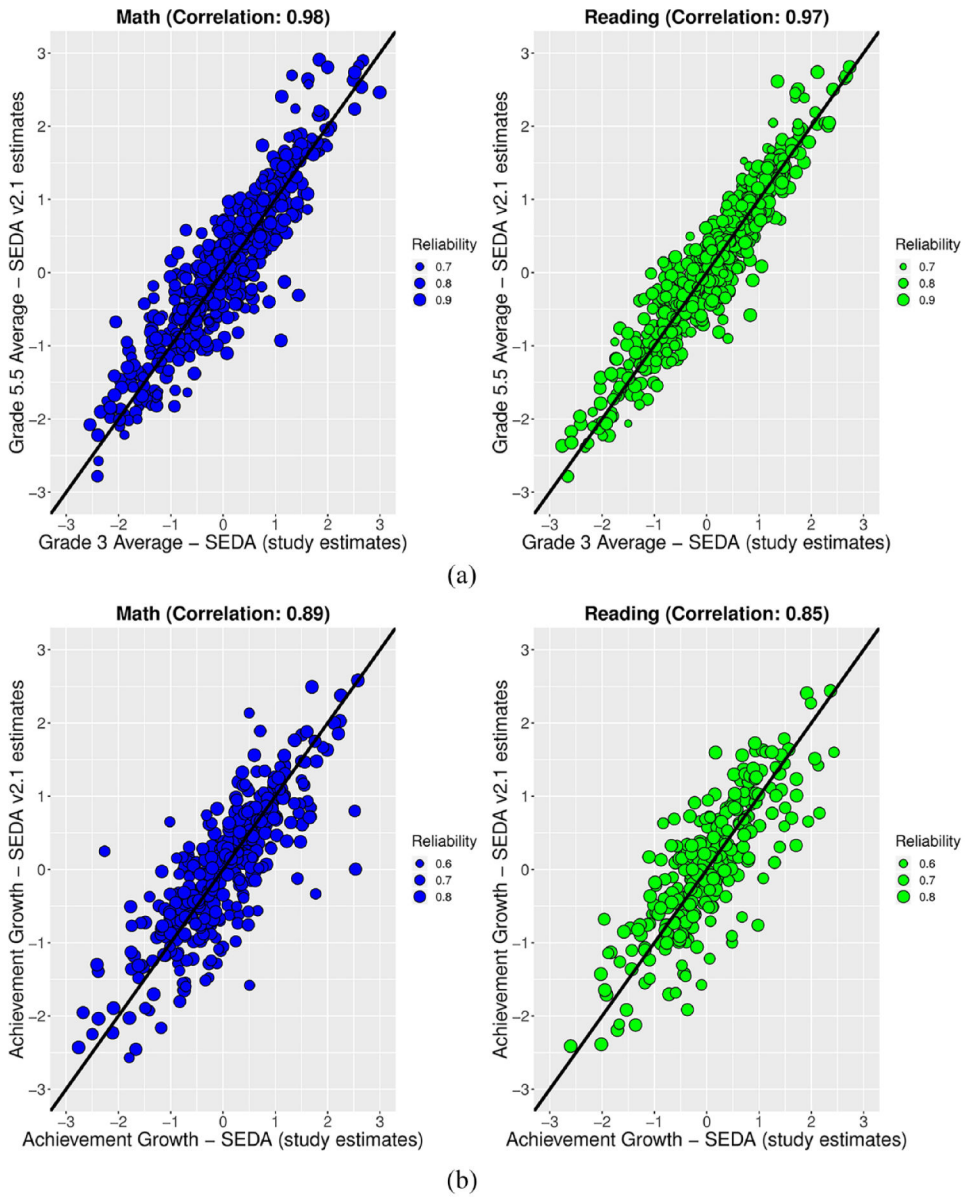
FIGURE A1. Comparison of the standardized "pooled" SEDA EB estimates used in this study with the standardized "pooled" SEDA v2.1 estimates (reported on Grade within Cohort Standardized (gcs) Scale). Note that the intercept parameter from the SEDA v2.1 estimates is the average Grade 5.5 score for cohort 2006.5, whereas the intercept from our model is the average Grade 3 score for cohort 2004.5. Disattentuated correlations are reported in the figure titles.

(a) SEDA EB estimates of achievement status.

(b) SEDA EB estimates of achievement growth.

FIGURE A2. Comparison of the MAP Growth EB estimates (from the quadratic model) with the standardized "pooled" SEDA v2.1 estimates (reported on Grade within Cohort Standardized (gcs) Scale). Note that the intercept parameter from the SEDA v2.1 estimates is the average Grade 5.5 score for cohort 2006.5, whereas the intercept from the MAP Growth model is the average Grade 3 score for cohort 2004.5. Disattentuated correlations are reported in the figure titles.

(a) EB estimates of achievement status.
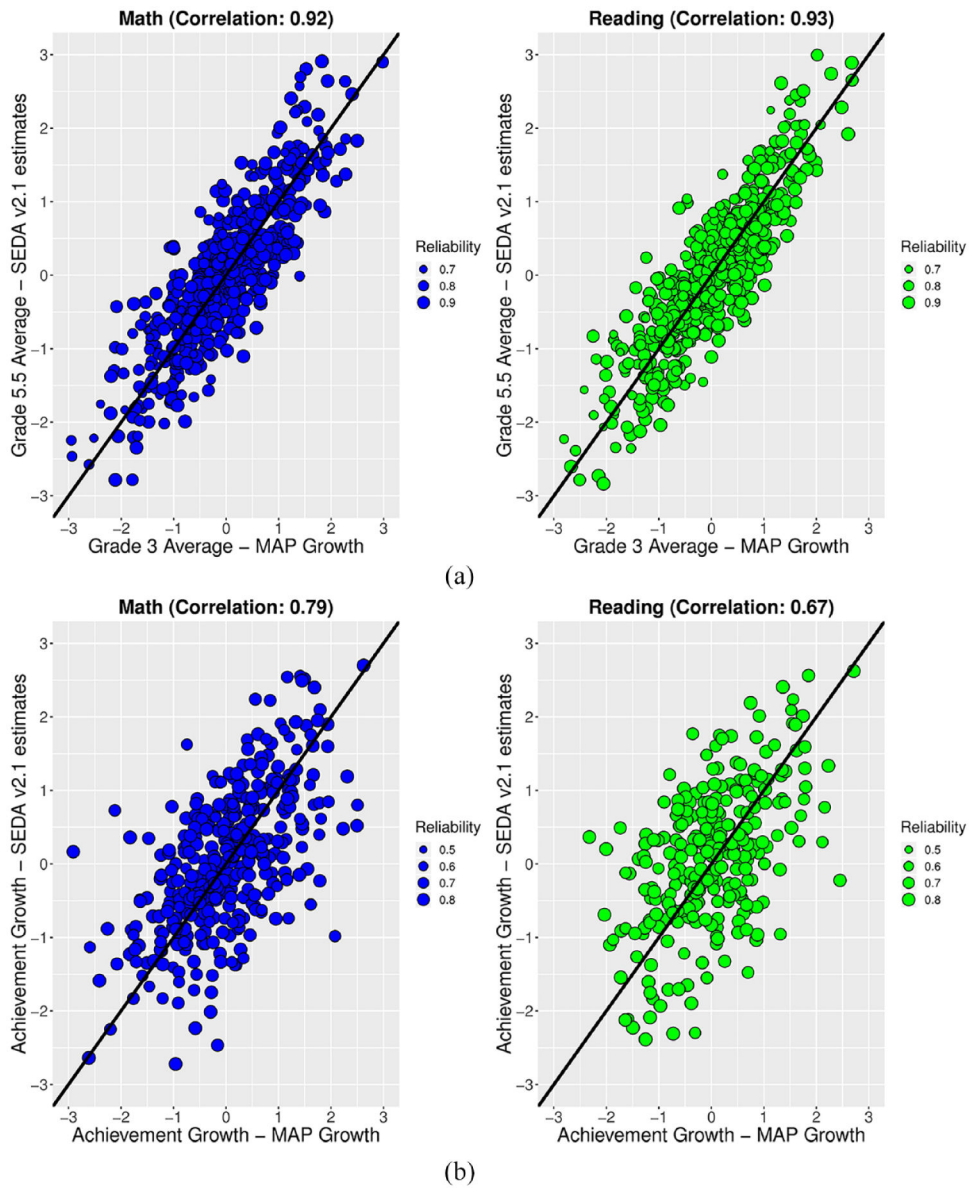
(b) EB estimates of achievement growth.

Authors

MEGAN KUHFELD is a research scientist II for the Collaborative for Student Growth at NWEA. Her research seeks to understand students' trajectories of academic and social-

emotional learning and the school and neighborhood influences that promote optimal growth.

THURSTON DOMINA is an associate professor of educational policy and sociology in the School of Education at the University of North Carolina at Chapel Hill. He works in partnership with educational practitioners to better understand the relationship between education and social inequality in the contemporary United States.

PAUL HANSELMAN is an assistant professor of sociology at the University of California, Irvine. His research focuses on educational inequalities in schools—where they come from and what we can do about them, with particular interest in understanding how local contexts shape educators' and students' experiences and their implications for economic, racial, and gender disparities.

References

Badger E, & Quealy K (2017, 12 5). How effective is your school district? A new measure shows where students learn the most. The New York Times. Retrieved from https://www.nytimes.com/interactive/2017/12/05/upshot/a-better-way-to-compare-public-schools.html

Bloom HS, Hill CJ, Black AR, & Lipsey MW (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. Journal of Research on Educational Effectiveness, 1(4), 289–328.

Briggs DC (2013). Measuring growth with vertical scales. Journal of Educational Measurement, 50(2), 204–226.

Camera L (2018, 4 10). NAEP shows little to no gains in math, reading for U.S. students. US News and World Report. Retrieved from https://www.usnews.com/news/best-states/articles/2018-04-10/naep-shows-little-to-no-gains-in-math-reading-for-us-students

Campbell JR, Hombo CM, & Mazzeo J (2000). NAEP 1999 trends in academic progress: Three decades of student performance. Washington, DC: National Center for Education Statistics.

Dadey N, & Briggs DC (2012). A meta-analysis of growth trends from vertically scaled assessments. Practical Assessment, Research & Evaluation, 17(14). Retrieved from http://pareonline.net/getvn.asp?v=17&n=14

Domingue B (2014). Evaluating the equal-interval hypothesis with test score scales. Psychometrika, 79(1), 1–19. [PubMed: 24532164]

Fahle E, & Reardon SF (2018). How much do test scores vary among school districts? New estimates using population data, 2009–2015. Educational Researcher, 47(4), 221–234, doi:10.3102/0013189X18759524

Fahle E, Shear BR, Kalogrides D, Reardon SF, DiSalvo R, & Ho AD (2018). Stanford Education Data Archive Technical Documentation (Version 2.1). Retrieved from https://stacks.stanford.edu/file/druid:db586ns4974/SEDA_documentation_v21.pdf

Karabatos G (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. Journal of Applied Measurement, 2, 389–423. [PubMed: 12011506]

Miller CC, & Quealy K (2018, 6 13). Where boys outperform girls in math: Rich, White and suburban districts. The New York Times. Retrieved from https://www.nytimes.com/interactive/2018/06/13/upshot/boys-girls-math-reading-tests.html

National Assessment of Educational Progress. (2018). Interpreting NAEP reading results. Retrieved from https://nces.ed.gov/nationsreportcard/reading/interpret_results.aspx

NWEA. (2018, 3 14). NWEA and FEV tutor partner to give each student access to customized real-time tutoring and on-demand support. Retrieved from https://www.nwea.org/content/uploads/2018/03/NWEA_FEV_March-2018.pdf

Reardon SF (2018). Educational opportunity in early and middle childhood: Variation by place and age. Retrieved from https://cepa.stanford.edu/content/educational-opportunity-early-and-middle-childhood-variation-place-and-age

Reardon SF, Ho AD, Shear BR, Fahle E, Kalogrides D, & DiSalvo R (2018). Stanford Education Data Archive (Version 2.1). Retrieved from http://purl.stanford.edu/db586ns4974

Reardon SF, Kalogrides D, & Ho A (2018). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale (CEPA Working Paper No.16–09). Retrieved from https://cepa.stanford.edu/sites/default/files/wp16-09-v201807.pdf

Reardon SF, Kalogrides D, & Shores K (2018). The geography of racial/ethnic test score gaps (CEPA Working Paper No.16–10). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp16-10

Reardon SF, Shear BR, Castellano KE, & Ho AD (2017). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. Journal of Educational and Behavioral Statistics, 42, 3–45. doi:10.3102/1076998616666279

Shores K, & Steinberg MP (2017). The impact of the great recession on student achievement: Evidence from population data (CEPA Working Paper No.17–09). Retrieved from http://cepa.stanford.edu/wp17-09

Shores K, & Steinberg M (2018). The great recession, fiscal federalism and the consequences for cross-district spending inequality (CEPA Working Paper No.18–14). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp18-14.

Thissen D (2012). Validity issues involved in cross-grade statements about NAEP results. Washington, DC: American Institutes for Research, NAEP Validity Studies Panel.

Thum YM (2018, 7). The vertical, interval scale assumption of a computerized adaptive test: Evidence from additive conjoint measurement. Paper presented at the ITC Meetings, Montreal, Canada.

Thum YM, & Hauser CH (2015). NWEA 2015 MAP norms for student and school achievement status and growth. NWEA Research Report. Portland, OR: NWEA.

von Hippel PT, Workman J, & Downey DB (2018). Inequality in reading and math skills forms mainly before kindergarten: A replication, and partial correction, of "Are schools the great equalizer?" Sociology of Education, 91(4), 323–357.
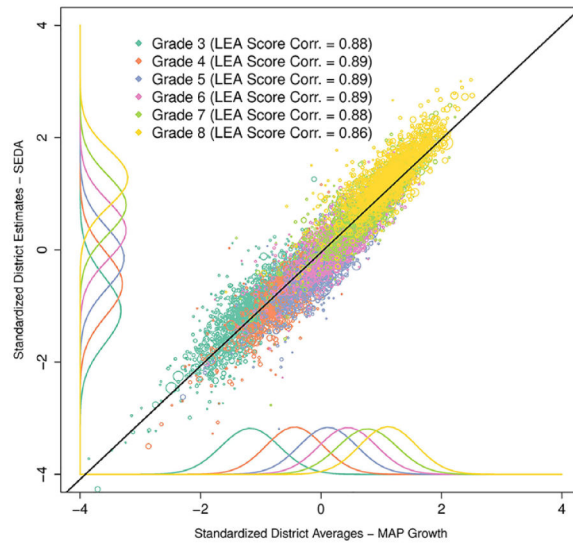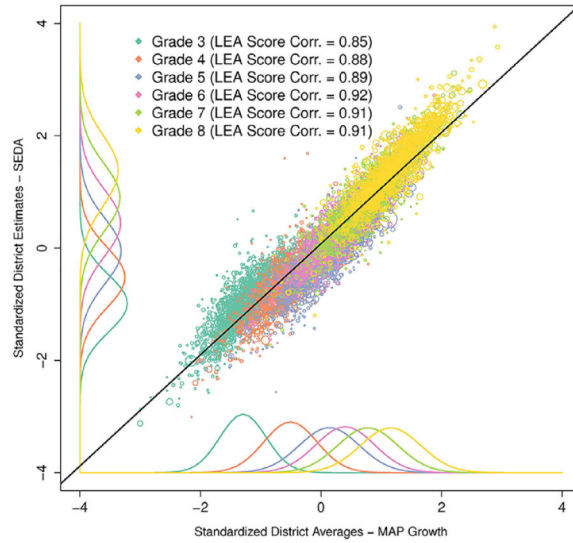
FIGURE 1.

Comparison of Stanford Educational Data Archive (SEDA) and MAP Growth district mean scores distributions by grade and subject in 2011–12 school year. Bubble size corresponds to the number of students testing in the district-grade-subject year combination. (a) English language arts scores. (b) Math scores.
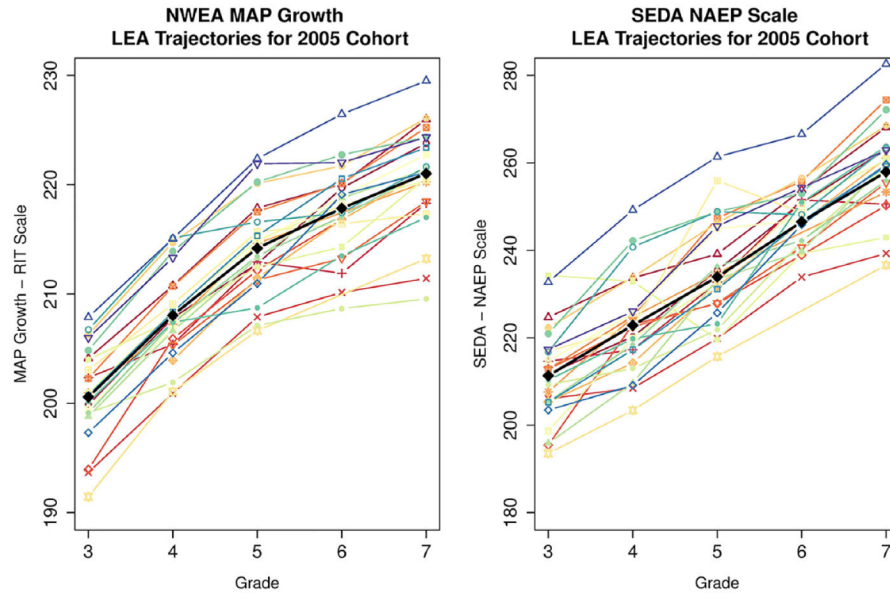
FIGURE 2.

Comparison of district ELA score trajectories for MAP Growth (RIT scale, left panel) and SEDA (NAEP scale, right panel) for a randomly selected set of 25 districts. The black line represents the average trajectory across the two cohorts, while the colored lines represent individual district trajectories. The students starting third grade in 2008 (e.g., the 2005 kindergarten cohort) are followed from third to seventh grades within the years of the study (2008–2009 to 2012–2013). Patterns vary somewhat depending on the set of districts that are randomly sampled. ELA = English language arts; SEDA = Stanford Educational Data Archive; NAEP = National Assessment of Educational Progress.
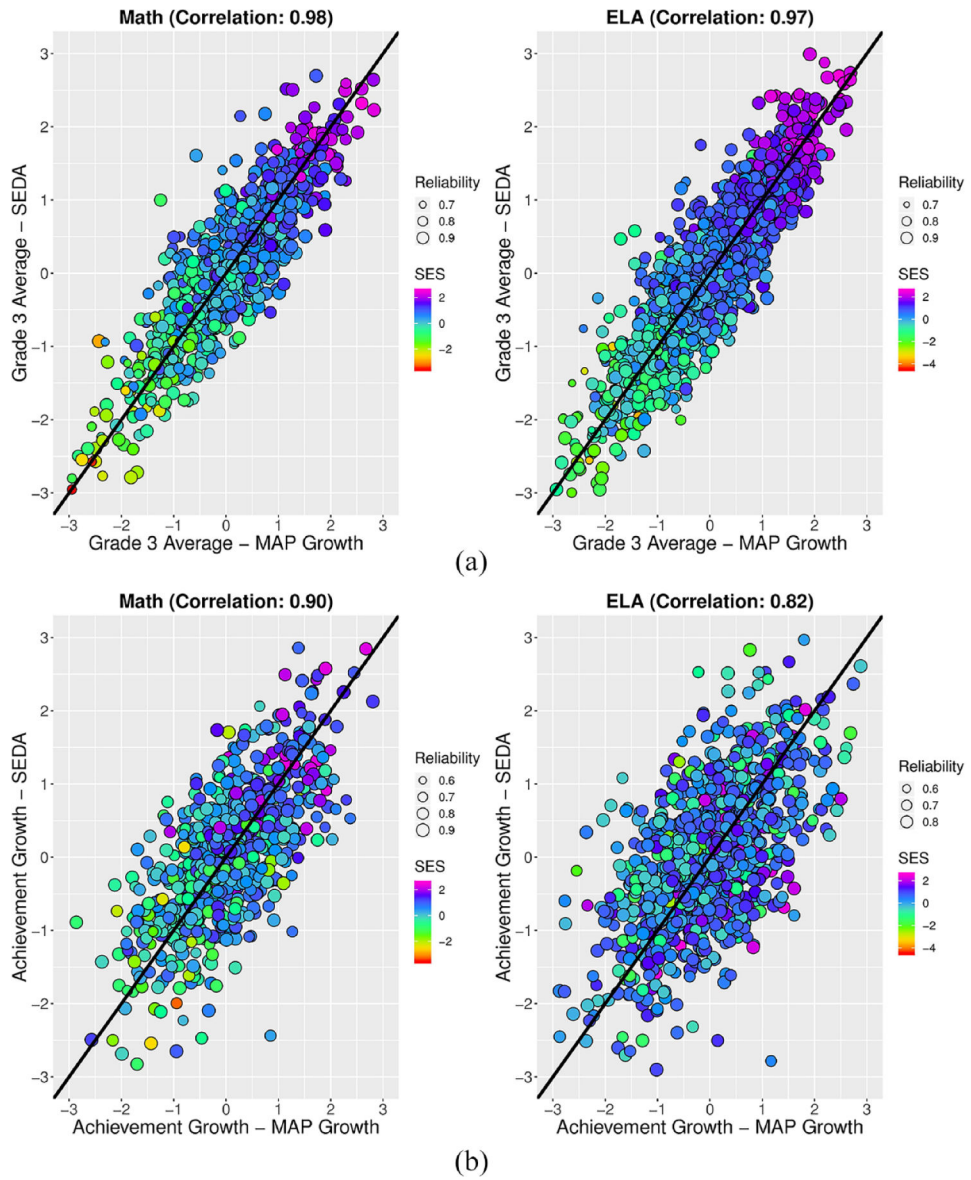
FIGURE 3.
Comparison of the outputs from the pooling model for SEDA and MAP Growth. The EB estimates shown are reported in a standardized metric. Bubble size corresponds to the reliability of the EB estimates and the color coding shows the district SES level. (a) EB estimates of the grade 3 average. (b) EB estimates of achievement growth. SEDA = Stanford Educational Data Archive; EB = empirical Bayes.

TABLE 1

Counts of Districts From SEDA and MAP Growth

Grade	SEDA districts					Matched analytic sample				
	2008–2009	2009–2010	2010–2011	2011–2012	2012–2013	2008–2009	2009–2010	2010–2011	2011–2012	2012–2013
3	10,547	10,658	10,643	10,659	10,638	763	907	1,046	1,228	1,449
4	10,579	10,655	10,705	10,702	10,635	780	926	1,082	1,238	1,442
5	10,603	10,687	10,722	10,783	10,670	793	939	1,067	1,253	1,444
6	10,623	10,709	10,762	10,755	10,680	782	924	1,066	1,215	1,409
7	10,530	10,607	10,658	10,684	10,529	735	860	984	1,122	1,327
8	10,576	10,666	10,659	10,667	10,518	665	813	937	1,058	1,245

Note. The number of SEDA districts represents the count of districts with a reported English language arts estimates. The matched analytic sample is the number of districts that tested in MAP Growth with 90% to 110% of the students tested in SEDA and are the set of districts used in subsequent analyses. For reference, there are approximately 14,372 geographic school districts according to SEDA, including many high school districts that would not have scores to contribute to these analyses. SEDA = Stanford Educational Data Archive.

TABLE 2

Demographic Characteristics of Full Set of SEDA Districts and Analytical Sample

Variable	Grade	2008–2009		2011–2012	
		Districts with SEDA scores	Districts with SEDA and MAP Growth scores	Districts with SEDA scores	Districts with SEDA and MAP Growth scores
N of districts	3	10,547	763	10,638	1,449
Average N of students	3	2,663.41	1,743.59	2,683.38	1,721.58
Percentage urban	3	0.07	0.05	0.07	0.05
SES	3	0.09	0.42	0.10	0.33
Percentage White	3	0.74	0.81	0.73	0.79
Percentage FRPL	3	0.36	0.28	0.42	0.37
Percentage ELL	3	0.04	0.04	0.05	0.04
Percentage SPED	3	0.13	0.14	0.14	0.14
N of districts	4	10,547	763	10,638	1,449
Average N of students	4	2,658.72	1,779.12	2,682.55	1,728.24
Percentage urban	4	0.07	0.05	0.07	0.05
SES	4	0.09	0.42	0.11	0.34
Percentage White	4	0.74	0.81	0.73	0.79
Percentage FRPL	4	0.35	0.28	0.42	0.37
Percentage ELL	4	0.04	0.04	0.05	0.04
Percentage SPED	4	0.13	0.14	0.14	0.14
N of districts	5	10,603	793	10,670	1,444
Average N of students	5	2,654.9	1,763.86	2,679.46	1,742.23
Percentage urban	5	0.07	0.05	0.07	0.04
SES	5	0.09	0.44	0.10	0.32
Percentage White	5	0.75	0.82	0.73	0.79
Percentage FRPL	5	0.35	0.27	0.42	0.37
Percentage ELL	5	0.04	0.04	0.05	0.04
Percentage SPED	5	0.13	0.14	0.14	0.14
N of districts	6	10,623	782	10,680	1,409
Average N of students	6	2,651.36	1,752.05	2,663.46	1,697.33
Percentage urban	6	0.07	0.05	0.07	0.04

Variable	Grade	2008–2009		2011–2012	
		Districts with SEDA scores	Districts with SEDA and MAP Growth scores	Districts with SEDA scores	Districts with SEDA and MAP Growth scores
SES	6	0.09	0.42	0.10	0.33
Percentage White	6	0.75	0.82	0.73	0.80
Percentage FRPL	6	0.33	0.26	0.40	0.35
Percentage ELL	6	0.04	0.04	0.05	0.04
Percentage SPED	6	0.13	0.14	0.14	0.14
N of districts	7	10,530	735	10,529	1,327
Average N of students	7	2,674.51	1,807.75	2,694.45	1,757.79
Percentage urban	7	0.07	0.05	0.07	0.05
SES	7	0.08	0.41	0.10	0.33
Percentage White	7	0.75	0.82	0.73	0.80
Percentage FRPL	7	0.32	0.25	0.39	0.34
Percentage ELL	7	0.04	0.04	0.05	0.04
Percentage SPED	7	0.13	0.14	0.14	0.14
N of districts	8	10,576	665	10,518	1,245
Average N of students	8	2,665.35	1,870.88	2,692.48	1,650.3
Percentage urban	8	0.07	0.05	0.07	0.04
SES	8	0.08	0.42	0.10	0.33
Percentage White	8	0.75	0.83	0.74	0.81
Percentage FRPL	8	0.32	0.25	0.39	0.34
Percentage ELL	8	0.04	0.04	0.05	0.04
Percentage SPED	8	0.13	0.14	0.14	0.14

Note. Two years were selected for the comparison of the SEDA and analytic sample, but the patterns are fairly consistent across years. While most of the reported variables are percentages than range from 0 to 1, SES is a district-level standardized variable with mean 0 and standard deviation of 1. The SES measure is computed as the first principal component score of the six district measures: median income, percentage of adults ages 25 and older with a bachelor's degree or higher, poverty rate for households with children ages 5 to 17, SNAP receipt rate, single mother–headed household rate, and employment rate for adults ages 25 to 64. SEDA = Stanford Educational Data Archive; SES = socioeconomic status; FRPL = free or reduced-price lunch; ELL = English language learner; SPED = special education.

TABLE 3

Estimates From the Pooling Models for SEDA and MAP Growth

	SEDA (NAEP scale) estimates		MAP Growth (linear model) estimates		MAP Growth (quadratic model) estimates	
	Math	**ELA**	**Math**	**ELA**	**Math**	**ELA**
Grade 3 average						
Estimate	230.25	207.93	205.84	201.36	199.49	203.51
SD	9.99	12.22	4.86	4.60	4.84	4.59
Reliability	0.75	0.80	0.76	0.75	0.81	0.80
Linear growth, Grades 3–8						
Estimate	11.39	11.97	6.56	4.81	9.79	9.49
SD	2.07	1.64	1.16	0.75	0.95	0.61
Reliability	0.58	0.50	0.65	0.50	0.65	0.49
Quadratic Growth, Grades 3–8						
Fixed effect estimate	—	—	—	—	−0.53	−0.67
Correlations						
Corr(Grade 3, linear growth)	−0.03	−0.38	−0.04	−0.25	0.07	−0.22
Corr(Grade 3 math, Grade 3 ELA)	0.94		0.95		0.93	
Corr(Math linear growth, reading linear growth)	0.73		0.94		0.85	

Note. All models were estimated with 60,543 district-grade-year-subject scores nested within 1,895 districts. Although the NAEP and MAP Growth scale ranges appear similar, direct comparisons between the estimates on the two scales are not possible. All of the fixed effects in each model were statistically significant. SEDA = Stanford Educational Data Archive; ELA = English language arts; NAEP = National Assessment of Educational Progress.

TABLE 4

Correlations Between Empirical Bayes Estimates and District Characteristics

	Math					ELA				
	G3 (MAP) (1)	G3 (SEDA) (2)	Growth (MAP) (3)	Growth (SEDA) (4)	Growth discrepancy (5)	G3 (MAP) (6)	G3 (SEDA) (7)	Growth (MAP) (8)	Growth (SEDA) (9)	Growth discrepancy (10)
G3 (MAP)	1.00*					1.00*				
G3 (SEDA)	0.98*	1.00*				0.97*	1.00*			
Growth (MAP)	0.21*	0.28*	1.00*			-0.26*	-0.14*	1.00*		
Growth (SEDA)	0.29*	0.10*	0.90*	1.00*		-0.28*	-0.43*	0.82*	1.00*	
Growth discrepancy	-0.15*	0.20*	0.50*	-0.55*	1.00*	0.02	0.33*	0.63*	-0.64*	1.00*
Percentage White	0.44*	0.50*	0.10*	0.15*	-0.06*	0.48*	0.52*	-0.06*	-0.11*	0.07*
Percentage Hispanic	-0.24*	-0.30*	-0.05*	-0.07*	0.03	-0.28*	-0.31*	0.08*	0.13*	-0.05
Percentage Black	-0.32*	-0.35*	-0.13*	-0.15*	0.02	-0.30*	-0.35*	-0.03	0.02	-0.06*
SES	0.71*	0.70*	0.36*	0.35*	0.01	0.73*	0.73*	0.09*	-0.04	0.15*
Percentage ELL	-0.27*	-0.30*	-0.01	-0.06*	0.06*	-0.31*	-0.33*	0.11*	0.13*	-0.02
White-Black segregation	-0.09*	-0.11*	-0.05*	-0.08*	0.03	-0.10*	-0.12*	-0.01	-0.01	0.00
Gini coefficient	-0.40*	-0.42*	-0.21*	-0.22*	0.02	-0.42*	-0.44*	-0.02	0.03	-0.07*
90/10 income ratio	-0.32*	-0.32*	-0.13*	-0.14*	0.02	-0.32*	-0.32*	0.01	0.04	-0.04
Total enrollment	0.10*	0.04	0.05	0.00	0.05	0.08*	0.04	0.02	0.00	0.03
Pupil-teacher ratio	-0.06*	-0.08*	-0.06*	-0.04	0.00	-0.07*	-0.07*	-0.01	0.02	-0.03
Revenue per pupil	0.03	0.05*	0.13*	0.09*	0.03	0.02	0.06*	0.07*	0.02	0.07*

Note. ELA = English language arts; G3 = Grade 3; SES = socioeconomic status; Growth discrepancy = standardized (z-score) EB estimate of Growth (MAP) subtracted by the standardized (z-score) EB estimate of Growth (SEDA); SEDA = Stanford Educational Data Archive; EB = empirical Bayes.

*
$p < .05$.

TABLE 5

Correlations for State-by-State Comparison of SEDA and MAP Growth Estimates

State	Correlation of Grade 3 scores (MAP, SEDA)	Correlation of growth scores (MAP, SEDA)	N	Total N of geographic districts in state	MAP Growth M	MAP Growth SD	SEDA M	SEDA SD
ELA								
IA	0.73	0.58	82	362	−0.02	0.93	0.08	0.84
IL	0.96	0.66	139	873	−0.09	0.82	−0.13	0.84
IN	0.90	0.45	33	300	−0.01	0.68	−0.06	0.70
KS	0.61	0.60	54	307	−0.70	1.04	−0.26	1.34
KY	0.73	0.52	39	176	−0.74	0.87	−0.26	1.21
ME	0.88	0.58	40	250	0.25	0.88	0.04	0.97
MI	0.91	0.45	40	608	−0.25	0.89	0.22	0.76
MN	0.87	0.77	94	377	0.29	1.05	−0.16	1.02
MT	0.94	0.50	20	320	0.11	0.74	0.31	1.25
ND	0.62	0.71	37	187	−0.18	1.09	−0.35	1.18
NE	0.88	0.67	28	260	−0.01	1.11	0.39	1.25
NH	0.88	0.74	53	156	0.23	1.02	−0.10	1.17
NJ	0.93	0.25	21	529	0.40	0.91	0.50	0.80
SC	0.94	0.61	42	89	−0.12	0.76	−0.14	0.68
WA	0.90	0.61	54	303	0.60	0.99	0.37	1.16
WI	0.87	0.57	98	420	0.24	0.95	0.03	0.82
WY	0.90	0.51	20	58	0.13	0.74	0.54	0.87
Math								
IA	0.72	0.71	104	362	−0.09	1.01	−0.01	0.84
IL	0.95	0.80	166	873	−0.18	0.88	−0.07	0.90
IN	0.88	0.58	43	300	0.05	0.69	−0.02	0.79
KS	0.54	0.50	59	307	−0.26	0.83	−0.13	1.33
KY	0.58	0.74	66	176	−0.83	0.88	−0.60	1.15
ME	0.89	0.72	45	250	0.04	0.89	−0.04	0.81
MI	0.91	0.60	55	608	−0.61	0.93	−0.09	0.78
MN	0.80	0.79	102	377	0.62	1.04	0.15	1.21

State	Correlation of Grade 3 scores (MAP, SEDA)	Correlation of growth scores (MAP, SEDA)	N	Total N of geographic districts in state	MAP Growth		SEDA	
					M	SD	M	SD
MT	0.92	0.66	26	320	0.13	0.78	0.29	1.19
ND	0.65	0.71	41	187	0.14	1.10	−0.07	1.23
NE	0.87	0.78	42	260	0.23	1.12	0.13	1.15
NH	0.88	0.82	56	156	0.22	1.00	0.15	1.08
NJ	0.90	0.58	29	529	0.69	0.97	0.84	0.82
SC	0.91	0.76	45	89	−0.11	0.86	−0.45	0.82
WA	0.87	0.70	67	303	0.41	0.88	−0.08	1.00
WI	0.86	0.68	133	420	0.18	0.86	0.22	0.77
WY	0.88	0.64	22	58	0.35	0.78	0.54	0.69

Note. *N* = the number of districts with a reliable SEDA and MAP Growth estimates for both the Grade 3 and Grades 3 through 8 growth estimates. Only states with a minimum of 20 districts with both SEDA and MAP Growth estimates in a given subject are reported in this table. ELA = English language arts; SEDA = Stanford Educational Data Archive.