ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-214

The learning curve: Revisiting the assumption of linear growth across the school year

Megan Kuhfeld
NWEA

James Soland
University of Virginia

Important educational policy decisions, like whether to shorten or extend the school year, often require accurate estimates of how much students learn during the year. Yet, related research relies on a mostly untested assumption: that growth in achievement is linear throughout the entire school year. We examine this assumption using a data set containing math and reading test scores for over seven million students in kindergarten through 8th grade across the fall, winter, and spring of the 2016-17 school year. Our results indicate that assuming linear within-year growth is often not justified, particularly in reading. Implications for investments in extending the school year, summer learning loss, and racial/ethnic achievement gaps are discussed.

VERSION: April 2020

Suggested citation: Kuhfeld, Megan, and James Soland. (2020). The learning curve: Revisiting the assumption of linear growth across the school year. (EdWorkingPaper: 20-214). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/bvg0-8g17

The learning curve: Revisiting the assumption of linear growth across the school year

Megan Kuhfeld
Research Scientist, NWEA

James Soland
Assistant Professor, University of Virginia
Affiliated Research Fellow, NWEA

CONTACT INFORMATION:

Megan Kuhfeld
121 N.W. Everett Street
Portland, OR 97209
Ph. 503-548-5295
megan.kuhfeld@nwea.org

The learning curve: Revisiting the assumption of linear growth across the school year

Abstract

Important educational policy decisions, like whether to shorten or extend the school year, often require accurate estimates of how much students learn during the year. Yet, related research relies on a mostly untested assumption: that growth in achievement is linear throughout the entire school year. We examine this assumption using a dataset containing math and reading test scores for over seven million students in kindergarten through 8th grade across the fall, winter, and spring of the 2016-17 school year. Our results indicate that assuming linear within-year growth is often not justified, particularly in reading. Implications for investments in extending the school year, summer learning loss, and racial/ethnic achievement gaps are discussed.

Introduction

Understanding a range of topics in education practice and policy is dependent on having accurate estimates of how much students learn during the year. For example, research examining potential benefits of lengthening the school year, as well as how much achievement drops during the summer when students are out of school, require that we accurately capture how much students grow academically during the school year. Such studies often assume that students will continue learning at the same rate throughout the entire school year, including after standardized testing is administered in the spring. That is, because spring testing does not typically occur at the very end of the school year, research on how much achievement increases during the year and drops (on average) when students are not in school frequently uses linear extrapolation of test scores to account for in-school time after student testing (Downey, von Hippel, & Broh, 2004; Quinn, Cooc, McIntyre, Gomez, 2016; von Hippel, Workman, & Downey, 2018). The assumption of linear within-year growth is fundamental to much of what we know about the effects of extended learning time, summer loss, and whether educational inequality widens more during or outside of school time (e.g., Atteberry & McEachin, 2019; Downey, von Hippel, & Broh, 2004; von Hippel, Workman, & Downey, 2018).

Yet, little evidence exists on whether such an assumption is justified. This omission in the literature is surprising given, on a basic level, the claim of linear within-year growth may lack face validity. Is it plausible to argue that the same amount of learning occurs during the week immediately before end-of-grade testing as during the week immediately before summer break? While assumptions about the linearity of school-year growth have not gone entirely unexamined, we are only aware of two studies that have presented evidence on this topic explicitly (Fitzpatrick, Grissmer, & Hastedt, 2011; Hayes & Gershenson, 2016). Other studies have also briefly mentioned the issue, but none of them made testing the assumption of linear within-year growth their primary focus, nor did they present details on analyses checking that assumption (Downey, von Hippel, & Broh, 2004; Quinn et al., 2016; von Hippel & Hamrock, 2019).

One reason this issue has largely gone unexamined is the lack of available data. Most achievement tests are administered once per year in the spring. Yet, to examine assumptions about the functional form of within-year growth, researchers require not two but three timepoints per school year. Thus, researchers are often left with no choice but to assume the linearity of within-year growth. In this study, we test the assumption of within-year linearity using data from the Measures of Academic Progress (MAP) Growth tests in math and reading. MAP Growth is a progress-monitoring assessment with a vertical scale that is typically administered in fall, winter, and spring to over 9 million students annually (NWEA, 2020). These data allow us to fit a within-year polynomial for a huge sample of the country's elementary school children. In so doing, we can compare model fit between models that do and do not assume within-year growth is linear. We can also see if the assumption affects estimates relevant to extending the school year, quantifying summer loss, and estimating achievement gaps. Specifically, we investigate four research questions:

1. Is there evidence that within-year growth is linear across kindergarten to 8^{th} grade in math and reading?
2. Are inferences about extending the school year sensitive to the assumption of linear growth throughout the school year?
3. How sensitive are estimates of summer learning loss to the assumption of linearity?

4. To the degree we see patterns of nonlinear growth, are certain racial/ethnic and gender groups more likely to show a relative slowdown in learning rates across the school year that would contribute to achievement gaps?

Background

Evidence of linearity

Evidence confirming or disconfirming the tenability of the within-year linearity assumption is limited. Only a handful of studies even examine the issue, and all but two of the studies treat the issue as peripheral to the main research question of interest. First, using the Early Childhood Longitudinal Study - Kindergarten Cohort (ECLS-K), Fitzpatrick, Grissmer, and Hastedt (2011) and Hayes and Gershenson (2016) each exploited the quasi-randomness of test dates and provided some initial evidence that students' school year academic gains in kindergarten and 1st grade were best fit by a linear model. However, both studies relied on only two testing occasions (fall and spring) and therefore could only examine learning gains occurring during the very beginning or end of the school year. Furthermore, it is unclear whether evidence supporting linearity in kindergarten and 1st grade can be generalized to the upper elementary and middle school grades, when preparation for standardized testing greatly shapes the instructional organization of the school year. Second, von Hippel and Hamrock (2019) reported testing the linearity of within-year growth using NWEA's MAP Growth data collected between 2008 and 2010 in 419 schools, but did not describe the methods used or grades examined. In short, within-year linearity is a relatively untested assumption and, as we describe below, is also fundamental to important inferences about education.

Implications for extending the school year

One consideration often facing policymakers that relates to the nature of within-year learning trajectories is whether to shorten or lengthen the school year. In particular, the question of whether to extend the school year remains a hotly debated policy topic (Patall, Cooper, & Allen, 2010). Despite the debate, lengthening school calendars is a strategy widely used in practice. Rocha (2008) found that more than 300 initiatives to extend learning time were launched between 1991 and 2007 in high-poverty and high-minority schools in 30 states. One reason for the popularity of this approach may be the research base. According to a literature review conducted by Patall, Cooper, and Allen (2010), 15 empirical studies of various designs have been conducted on the topic since 1985, most of which used descriptive analyses rather than quasi-experimental or experimental designs. Those studies suggested that extending school time can be an effective way to support student learning, especially for students most at risk of dropping out and when a deliberate strategy is used to maximize the utility of that extended learning time.

While the studies reviewed by Patall, Cooper, and Allen (2010) suggest that extending the school year is promising, their descriptive nature means they often rely on assumptions about how growth occurs during the school year. That is, most evidence from such studies assumes that the benefits of extending the school year are linear such that every day increase is associated with some commensurate gain in achievement. Thus, policymakers wishing to apply those findings can roughly estimate how much learning might improve given an extension of a certain number of days. However, if within-year learning is nonlinear—and especially if those nonlinear trends differ by context—then the implications of the findings in the studies reviewed by Patall, Cooper, and Allen (2010) are much harder to identify. For example, if learning gains taper dramatically in a given district towards the end of the year, then investing in a longer school year may be less likely to improve achievement relative to other possible strategies. In summary, basing inferences about the benefits of extending the school year on studies that assume linear growth throughout the school year may lead to a misunderstanding of the actual benefit to students' academic achievement.

Implications for Studying Summer Learning Loss

Since students rarely are tested on the first and last day of the school year, quantifying summer learning loss typically involves the models to separate the effects of in-school time (including instruction that occurs after testing in the spring or before testing in the fall) on achievement from the effects of summer vacation. Isolating the learning occurring during the summer is routinely achieved through linearly extrapolating test scores to the start and end of the school year (i.e., summer loss models assume that learning occurs at the same linear rate before the first test of the school year and after the last test). In some studies (e.g., Quinn, 2015; Atteberry & McEachin, 2019), test scores were projected to the start and end date based on each student's individual growth rate prior to estimating summer loss. In other studies, school-year and summer learning patterns were estimated using a hierarchical linear growth model that assumes linear growth during the school year (Downey, von Hippel, & Broh, 2004; Quinn et al., 2016; von Hippel, Workman, & Downey, 2018, von Hippel & Hamrock, 2019). As von Hippel and colleagues (2018) note, this model "implicitly extrapolates beyond the test dates to the scores that would have been achieved on the first and last day of the school year" (p. 335).

As noted earlier, many of the summer learning loss studies included footnotes and/or supplemental materials addressing checks for non-linearity with the ECLS-K data (Downey, von Hippel, & Broh, 2004, pg. 620; Quinn, Cooc, McIntyre, & Gomez, 2016, pg. 452) and NWEA data (von Hippel & Hamrock, 2019, pg. 55; Atteberry & McEachin, 2019, pg. 40). Only

Atteberry and McEachin (2019) provided a comparison between the estimated degree of summer loss between the linearly projected scores and the observed fall and spring scores to investigate the degree to which projecting scores influences measurement of summer learning loss. In their study, the average student took the fall test about 26 days after the first day of school and took the spring test 39 days before the last day of school. When linearly projecting test scores to the start and end of the school year, they found that the 50th percentile student shows summer learning loss during each summer subsequent to 1st to 8th grade in both math and reading, whereas the observed test scores indicate that 50th percentile student was only losing ground in a subset of grades. These findings imply that the estimated impact of summer loss varies greatly depending on how test scores are projected to the start/end of the school year. However, as far as we are aware, no one has examined the sensitivity of summer learning drops to projecting test scores assuming students' growth trajectories are linear versus non-linear.

Implications for Studying Achievement Gaps

Given achievement gaps are associated with a host of long-term outcomes like postsecondary attainment and earnings (e.g., Neal, 2006), there is much interest in how achievement gaps develop over the course of students' schooling. Yet, the majority of these studies assume that achievement gaps growth linearly during the school year, which may not be the case. For example, Robinson and Lubienski (2011) showed that there is not a male-female mathematics gap in kindergarten using the ECLS-K, but that females lose some ground during elementary school before closing the gap again in middle school. Related studies show that mathematics gaps in kindergarten increased, on average, by 3rd grade (Husain & Millimet, 2009; LoGerfo, Nichols, & Chaplin, 2006; Rathbun, West, & Hausken, 2004). As for reading, Robinson and Lubienski (2011) also found that gaps are narrow across most grades, but widen over time among initially low-achieving students. For example, the ECLS-K achievement test showed a 0.2 SD gap at 4^{th} grade and more than 0.3 SD in 8^{th} grade and later (Robinson & Lubienski, 2011). Similarly, Husain and Millimet (2009) found that low-achieving males tend to lose ground in reading between kindergarten and 3^{rd} grade.

Similar evidence has been accrued for Black-White achievement gaps in studies that also tend to assume within-year growth is linear and, thus, that gaps widen linearly as well. For instance, gaps appeared to widen by 5^{th} grade, reaching 0.75 SD in reading and 1.0 SD in mathematics (Reardon & Galindo, 2009). A study using the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development showed that the mathematics gap narrowed between kindergarten and 3^{rd} grade, but widened in reading during that same period (Murnane et al., 2006). Another study using a different dataset suggested that mathematics and reading gaps grow between 1^{st} and 2^{nd} grade, then increase idiosyncratically in subsequent elementary grades (Phillips, Crouse, & Ralph, 1998). Trends in later grades are less clear, though gaps persist. A pair of studies using state data showed the mathematics gap growing from .59 to .70 SD between 3^{rd} and 8^{th} grade in Texas (Hanushek & Rivkin, 2006), but only growing from .77 to .81 SD in North Carolina (Clotfelter, Ladd, & Vigdor, 2006). Meanwhile, research using the National Assessment of Education Progress (NAEP) long-term trends data found that the mathematics gap widens between ages 9 and 13 (Ferguson, 2001; Neal, 2006; Phillips, Crouse, & Ralph, 1998).

A growing literature also considers whether inequality by subgroup grows during the school year on average across schools, oftentimes assuming linear within-year growth in the process. Perhaps most prominently, Downey, von Hippel, and Broh (2004) found that schools serve as important equalizers with most gaps growing faster during summer than during the school year (the Black-White gap being the one exception). This issue has been revisited many times over the intervening years. von Hippel, Workman, and Downey (2018) similarly showed that socioeconomic gaps often shrink during the school year and expand during the summer, while the Black-White gap tends to grow during school. However, other work, including research by von Hippel and Hamrock (2019), found that different gaps expand during different seasons and to varying degrees depending on data sources and measures of achievement.

Despite the attention paid to how gaps develop as student's move through school, there is virtually no evidence on whether these gaps develop linearly or nonlinearly within school years. That is, we do not know if gaps widen at a consistent rate over the course of a school year or if they widen more during particular times in the year like just before summer break. Once again, this omission in the literature likely occurs due in part to the limited availability of data from test scores administered three or more times per year.

Data and Measures

Analytic Sample

The data for this study are from the Growth Research Database (GRD) at NWEA. School districts partner with NWEA to monitor elementary and secondary students' reading and math growth throughout the school year, with assessments typically administered in the fall, winter, and spring. We use the test scores of over seven million kindergarten to 8th grade students in 16,824 schools from the 2017-18 school year. Sixty-seven percent of students in our sample are assessed during all three terms, with the remaining 33% assessed during only one or two of the terms. The GRD also includes demographic information, including student race/ethnicity, gender, and age at assessment, though student-level socioeconomic status is not available. Table 1 provides descriptive statistics for the sample by subject and grade. In each grade, we observe between 500,000 and 800,000 students. Overall, the sample is 51% male, 48% White, 17% Black, 4% Asian, and 18% Hispanic.

The set of schools partnering with NWEA to administer MAP Growth in 2017-18 represents approximately one in four US public schools. While extremely large, the sample consists of schools that partner with NWEA of their own volition and is therefore not inherently nationally representative. Table 2 provides a comparison of the school characteristics of our sample of 16,824 schools with the population[1] of 62,601 US public schools serving grades K-8 based on school-level data from the 2017-18 Common Core of Data (CCD). From these data, we are able to characterize our sample of schools based on a number of characteristics, including percentage of students eligible for free or reduced price lunch (FRPL), locale (e.g., urban vs. rural), enrollment by grade, and distribution of racial/ethnic groups. Overall, the NWEA sample of schools closely matches the national distribution of schools in urbanicity, percentage of FRPL eligibility, and percentage of White students, though schools serving a high percentage of Hispanic students are slightly underrepresented.

Measures of Achievement

Student test scores from NWEA's MAP Growth reading and math assessments are used in this study. MAP Growth is a computer adaptive test—which means measurement is precise even for students above or below grade level—and is vertically scaled to allow for the estimation of gains across time. Each test begins with a question appropriate for the student's achievement level (either based on a student's past performance or grade-level expectations), and then adapts throughout the test in response to student performance. The MAP Growth assessments are typically administered three times a year (fall, winter, and spring) and are aligned to state content

[1] We define the population of schools as the set of US public schools in the 50 states plus the District of Columbia that reported enrollment of at least one K-8 student during the 2017-18 school year within the CCD data file.

standards. Test scores are reported on the RIT (Rasch unIT) scale, which is a linear transformation of the logit scale units from the Rasch item response theory model. Table 3 provides the mean and standard deviations (SD) of students' RIT scores by term, grade, and subject.

Months of School Exposure

Schools using MAP Growth assessments set their own testing schedules, resulting in considerable variation around when students take the tests during the school year. To account for time in school before testing, we draw on district calendars for the 2017-18 school year from participating school districts. We calculate "months of exposure" to school at each test event as the total number of days elapsed between the school's start date and testing date(s) divided by 30. For example, a hypothetical student who started school on the first of August and tested three times (say September 5th, January 10th, and April 15th) would have a vector of months of exposure values of 1.17, 5.40, and 8.57 months. Table 3 provides the mean and SD of students' observed months of school by term, grade, and subject. The average student in our sample takes the fall test three weeks into the school year and the spring test with slightly over one month of school remaining.

Analyses

Descriptive Analyses. We first look at the variation in testing dates within MAP Growth data. Panel A of Figure 1 presents the frequency of test events within each month during the 2017-18 school year. Fall testing peaks between August and September, winter testing primarily takes place in December and January, and spring testing primarily occurs in April and May, though there is a fair amount of variability overall. However, our primary unit of time is not the date of testing but months of exposure to school, which accounts for differences in school start date. Panel B in Figure 1 in the supplemental materials displays the distribution of months of exposure.

To help answer the first research question, for each pair of terms (fall-winter and winter-spring), we first estimate the standardized mean difference effect sizes between terms. The means and SDs that we use in the effect size calculation are estimated pooling all students within a term (fall, winter, or spring), which ignores differences in testing month within a term. The standardized gain[2] between fall and winter test scores is

$$\frac{\overline{\mathrm{RIT}}_{Wg}-\overline{\mathrm{RIT}}_{Fg}}{\sqrt{\frac{(\mathrm{N}_{\mathrm{Wg}}-1)\mathrm{SD}_{Wg}^2+(\mathrm{N}_{Fg}-1)\mathrm{SD}_{Fg}^2}{\mathrm{N}_{Wg}+\mathrm{N}_{\mathrm{Fg}}-2}}}, \quad (1)$$

where $\overline{\mathrm{RIT}}_{Wg}$ is the average winter test score in grade g, $\overline{\mathrm{RIT}}_{Fg}$ is the average fall test score in grade g, SD_{Wg} and SD_{Fg} are the SD estimates in the winter and fall of grade g, and N_{Wg} and N_{Fg} are the observed sample size in the winter and fall of grade g respectively. The mean and SD estimates used in these calculations are all reported in Table 3. In addition, given the variation in testing windows observed with the MAP Growth data, we extend the above effect size equation to account for individual differences in the amount of time passed between two test events. Specifically, we calculate the average monthly gain between the fall and winter as

$$\frac{\sum_{i=1}^{N_g}\frac{\mathrm{RIT}_{Wi}-\mathrm{RIT}_{Fi}}{\mathrm{Mon}_{Wi}-\mathrm{Mon}_{Fi}}}{N_g}, \quad (2)$$

where RIT_{Wi} is student *i*'s winter test score, RIT_{Fi} is student "s fall test score, Mon_{Wi} is the month of exposure by the winter test event, Mon_{Fi} is the month of exposure by the fall test event, and N_g is the number of unique students with a fall and winter test score observed in grade g. Winter-spring effect sizes and average gains are calculated in the same manner.

[2] We also calculated standardized gains estimates accounting for within-person correlations across terms and results were highly similar.

Multilevel Growth Models. We also examine the assumption of linear school-year learning rates using a set of multilevel growth models (Raudenbush & Bryk, 2002). By coding time as months of exposure to school, our modeling approach adjusts for variation in the test dates and school-year start/end dates. We directly account for the clustering of students in schools by estimating three-level models, where MAP Growth test scores across fall, winter, and spring (level 1) are nested within students (level 2) and schools (level 3). For each grade and subject, three growth models are estimated.

Linear Growth Model. In the linear growth model, the test score y_{tij} for student *i* in school *j* at timepoint *t* is modeled as a linear function of the months (Months_{tij}) that the student has been in school:

$$\mathrm{y}_{tij} = \pi_{0ij} + \pi_{1ij}\mathrm{Months}_{tij} + e_{tij}. \quad (3)$$

At level 2 and 3, we include student- (r_{0ij} and r_{1ij}) and school-level (u_{00j} and u_{10j}) random effects for both the intercept and linear growth terms:

Level-2 Model (student (i) within school (j)): (4)

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

Level-3 Model (school (j)):

$$\beta_{00j} = \gamma_{000} + u_{00j}$$
$$\beta_{10j} = \gamma_{100} + u_{10j}$$

In this model, γ_{000} is the expected average test score on the first day of school and γ_{100} is the average linear gain in RIT points per month of school during the school year. An important assumption of this model is that growth is constant with respect to time, which implies that one month at the beginning of the school year is associated with the same amount of learning gains as one month at the end of the school year. The next two models discussed relax this assumption.

Quadratic Growth Model. The quadratic growth model expands upon the linear growth model by including a quadratic growth term at level 1:

$$\mathrm{y}_{tij} = \pi_{0ij} + \pi_{1ij}\mathrm{Months}_{tij} + \pi_{2ij}\mathrm{Months}_{tij}^2 + e_{tij}, \quad (5)$$

where Months_{tij}^2 is the squared number of months in school at timepoint *t.* After testing the model with random effects included for each parameter at the student-level, we determined that there are minimal within-school variations in the quadratic growth term, and so student-level random effects are only included for the intercept and linear growth term. At levels 2 and 3, the quadratic model is specified as:

Level-2 Model (student (i) within school (j)): (6)

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$
$$\pi_{2ij} = \beta_{20j}$$

Level-3 Model (school (j)):

$$\beta_{00j} = \gamma_{000} + u_{00j}$$
$$\beta_{10j} = \gamma_{100} + u_{10j}$$
$$\beta_{20j} = \gamma_{200} + u_{20j}$$

In this model, γ_{000} is the average test score on the first day of school, γ_{100} is the average instantaneous rate of change at the start of the school year, and γ_{200} is the average rate of change of the linear growth term for a one-unit change in time (e.g., the acceleration or deceleration in growth).

Piecewise Growth Model. In the piecewise model (also referred to as a linear spline model), we approximate a nonlinear function by tying together two linear growth components. The knot (transition point) is set at halfway through the school year (4.75 months) so that we can separately estimate linear growth rates during the first half and second half of the school year. The level-1 model is specified as:

$$\mathrm{y}_{tij} = \pi_{0ij} + \pi_{1ij}\text{Fall\_Months}_{tij} + \pi_{2ij}\text{Spring\_Months}_{tij} + e_{tij}, \quad (7)$$

where Fall_Months_{tij} is the coding of time for the first 4.75 months of the school year (fall to winter) and $\text{Spring\_Months}_{tij}$ is the coding of time for the second 4.75 months (winter to spring). The level-2 and level-3 parts of the model are specified in a similar manner to the other models:

Level-2 Model (student (i) within school (j)): (8)

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$
$$\pi_{2ij} = \beta_{20j} + r_{2ij}$$

Level-3 Model (school (j)):

$$\beta_{00j} = \gamma_{000} + u_{00j}$$
$$\beta_{10j} = \gamma_{100} + u_{10j}$$
$$\beta_{20j} = \gamma_{200} + u_{20j}$$

Model Fit Assessment. All models reported in this study are estimated using full information maximum likelihood with HLM version 7 (Raudenbush, Bryk, Cheong, & Congdon, 2011). To evaluate whether the quadratic model fits the data better than the linear model, we conduct a likelihood-ratio test (LRT) for each grade/subject. The LRT compares the deviance statistic (e.g., -2 times the log-likelihood estimate) of the more restricted linear model (nine estimated parameters) to the less two restricted models (13 estimated parameters for both the piecewise and quadratic model). The difference between the deviance estimate of the linear (D_0) and the alternative (D_1) model is chi-squared distributed with degrees of freedom equal to the difference in number of freed parameters between the two models (four in this case):

$$(D_0 - D_1) \sim \chi^2 \ (df = 4). \quad (9)$$

A significant LRT indicates that the inclusion of additional parameters improves model fit, while a non-significant LRT indicates that the additional parameters have not significantly improved model fit (e.g., the more restricted model is supported).

Extending the School Year. To examine the expected benefits from extending the school year (for the purpose of specificity, we examine the effect of a one month extension), we use the results from the multilevel growth models to predict each student's model-based learning gains across two school year lengths: (a) a 9-month school year and (b) a 10-month school year. For student *i* in school *j* in a given subject/grade, the expected gain during the school year under the linear model ($\widehat{\Delta}_{\text{ij,lin}}$) is predicted as

$$\widehat{\Delta}_{\text{ij,lin}} = \left(\hat{\gamma}_{100} + \hat{u}_{10j} + \hat{r}_{1ij}\right) * \text{Time}, \quad (10)$$

where $\hat{\gamma}_{100}$ is the estimated linear growth fixed effect from the multilevel growth model, $\hat{u}_{10j}$ is a school-level empirical Bayes (EB) estimate of the linear growth random effect, and $\hat{r}_{1ij}$ is a student-level EB estimate of the linear growth random effect. Time represents the months of school exposure and is set to either nine months (traditional schedule) or 10 months (extended school year). Under the quadratic model, the expected gain is

$$\widehat{\Delta}_{\text{ij,quad}} = \left(\hat{\gamma}_{100} + \hat{u}_{10j} + \hat{r}_{1ij}\right) * \text{Time} + \left(\hat{\gamma}_{200} + \hat{u}_{20j} + \hat{r}_{2ij}\right) * \text{Time}^2. \quad (11)$$

In this equation, $\hat{\gamma}_{200}$ is the estimated quadratic growth fixed effect, $\hat{u}_{20j}$ is a school-level quadratic growth EB estimate, and $\hat{r}_{2ij}$ is student-level quadratic growth EB estimate. The HLM software produces the EB estimates in the level-2 residual file for each model/grade/subject. Once the model-based gains are estimated for each individual, we calculate the average gain in

RIT units for a subject/grade for each model (linear or quadratic) and exposure to time (9 or 10 months):

$$\hat{\Delta}_{\text{lin,9mo}} = \frac{\sum_{i=1}^{N_g} \hat{\Delta}_{\text{ij,lin,9mo}}}{N_g}; \quad \hat{\Delta}_{\text{lin,10mo}} = \frac{\sum_{i=1}^{N_g} \hat{\Delta}_{\text{ij,lin,10mo}}}{N_g} \tag{12}$$
$$\hat{\Delta}_{\text{quad,9mo}} = \frac{\sum_{i=1}^{N_g} \hat{\Delta}_{\text{ij,quad,9mo}}}{N_g}; \quad \hat{\Delta}_{\text{quad,10mo}} = \frac{\sum_{i=1}^{N_g} \hat{\Delta}_{\text{ij,quad,10mo}}}{N_g}$$

where N_g is the total number of students in grade g. To estimate the extra learning gains from an additional month at the end of school year, we subtract the average expected gain across nine months from the average expected gain across 10 months for both the linear model ($\hat{\Delta}_{\text{lin,10mo}} - \hat{\Delta}_{\text{lin,9mo}}$) and the quadratic model ($\hat{\Delta}_{\text{quad,10mo}} - \hat{\Delta}_{\text{quad,9mo}}$). Finally, we standardize each set of estimates by the pooled SD for the corresponding grade/subject (estimated using the fall and spring SD estimates reported in Table 3). The standardized gain from an additional month of schooling assuming a linear model is therefore reported as:

$$\frac{\hat{\Delta}_{\text{lin,10mo}} - \hat{\Delta}_{\text{lin,9mo}}}{\sqrt{\frac{(\text{N}_{Fg}-1)\text{SD}_{Fg}^2 + (\text{N}_{Sg}-1)\text{SD}_{Sg}^2}{\text{N}_{Fg}+\text{N}_{Sg}-2}}}. \tag{13}$$

Summer Learning Loss. As with the previous research question, we use the EB estimates of students' linear and quadratic growth trajectories to project students' spring 2018 test score to the end of the school year. However, unlike in the previous analysis, we use each student's unique testing date and observed school calendar information to produce individual projections based on how many days remain in the student's school year. Since summer learning estimates require a spring and a fall test score, we estimate the summer growth for the subset of students who have observed fall test scores in the 2018-19 school year. We were able to locate and merge in fall 2018 test scores for 1st to 8th grade students for 71% of the original sample of students (unfortunately we are unable to follow our 8th grade cohort into 9th grade). Projections are conducted using both the linear and quadratic growth model within each grade and subject for both fall and spring scores (see Appendix B in the supplemental materials for more details), and then summer loss is calculated by subtracting students' projected fall test score from the projected spring test score.

Racial/ethnic and Gender Differences in Growth Deceleration. Lastly, we investigate whether there are group differences in the degree to which growth slows during the school year. Such estimates are relevant to understanding how achievement gaps change over the course of the school year. Specifically, we estimate a conditional quadratic growth model where the intercept, linear, and quadratic slope term are all regressed on fixed effects for student race/ethnicity and gender. That is, the coefficient for the interaction between the quadratic term and race/gender variable allows us to investigate whether there is an association between learning acceleration/deceleration and subgroup status. Additionally, to account for school-level socioeconomic status, we include controls for school percentage FRPL.

Level-1 Model (time (t) within student (i) within school (j)): (14)

$$\mathrm{y}_{tij} = \pi_{0ij} + \pi_{1ij}\mathrm{Months}_{tij} + \pi_{2ij}\mathrm{Months}_{tij}^2 + e_{tij},$$

Level-2 Model (student (i) within school (j)):

$$\begin{aligned}
\pi_{0ij} &= \beta_{00j} + \beta_{01j}\mathrm{Black}_{ij} + \beta_{02j}\mathrm{Hispanic}_{ij} + \beta_{03j}\mathrm{Asian}_{ij} \\
&\quad + \beta_{04j}\mathrm{OtherRace}_{ij} + \beta_{05j}\mathrm{Male}_{ij} + r_{0ij} \\
\pi_{1ij} &= \beta_{10j} + +\beta_{11j}\mathrm{Black}_{ij} + \beta_{12j}\mathrm{Hispanic}_{ij} + \beta_{13j}\mathrm{Asian}_{ij} \\
&\quad + \beta_{14j}\mathrm{OtherRace}_{ij} + \beta_{15j}\mathrm{Male}_{ij} + \quad r_{1ij} \\
\pi_{2ij} &= \beta_{20j} + \beta_{21j}\mathrm{Black}_{ij} + \beta_{22j}\mathrm{Hispanic}_{ij} + \beta_{23j}\mathrm{Asian}_{ij} \\
&\quad + \beta_{24j}\mathrm{OtherRace}_{ij} + \beta_{25j}\mathrm{Male}_{ij}
\end{aligned}$$

Level-3 Model (school (j)):

$$\begin{aligned}
\beta_{00j} &= \gamma_{000} + \gamma_{001}\%\mathrm{FRPL}_j + u_{00j} \\
\beta_{10j} &= \gamma_{100} + \gamma_{101}\%\mathrm{FRPL}_j + u_{10j} \\
\beta_{20j} &= \gamma_{200} + \gamma_{201}\%\mathrm{FRPL}_j + u_{20j}
\end{aligned}$$

Results

Is there evidence that within-year growth is linear across kindergarten to 8th grade in math and reading?

Table 4 presents fall-to-winter and winter-to-spring math and reading gains reported as standardized mean difference effect sizes. On average, fall and winter tests were 3.77 months apart and winter and spring tests were 3.95 months apart. Consistent with findings from other assessments (e.g., Dadey & Briggs, 2012; Bloom et al., 2008), the MAP Growth effect sizes are substantially larger in earlier grades than in middle school. The effect sizes show that, while the linear growth assumption appears tenable in some grades for math, growth in math for several grades and in reading across most grades appears to decelerate during the school year. The differences between the fall-winter and winter-spring effect sizes in math range from 0.17 SD in kindergarten to 0.01 SD in 4^{th} grade. In the middle school grades, fall-winter growth effect sizes in reading are approximately twice as large as the effect sizes observed between winter and spring. Table 4 provides the average monthly learning rates (accounting for time elapsed between test), which also consistently show higher growth during the fall-winter period.

Figure 2 displays the predicted average growth trajectories based on the linear, quadratic, and piecewise growth models. Within-year growth deceleration in reading is quite apparent for students in 4^{th} through 8^{th} grade. The quadratic parameter can be interpreted as the rate of change in the growth that occurs during the school year, with negative values indicating that growth is decelerating during the school year.

Table 5 shows the parameter estimates from the three unconditional multilevel growth models. The growth model results confirm the decelerating growth patterns seen in the standardized gain estimates. For example, whereas the estimate of the quadratic parameter (in

RIT units) is close to zero in math during 4th through 8th grade, the same parameter estimates in reading range from -0.04 RIT points per month (7th grade) to -0.10 RIT points per month (3rd grade). These coefficients appear small, but as demonstrated in Figure 2, this degree of growth deceleration leads to a noticeable tapering of growth in reading during the last two months of school. Further, in the piecewise model, the estimated linear growth in reading for the same grades from winter to spring is less than half the growth from fall to winter. Table A1 in the supplemental materials provides the sample size and deviance estimates for each multilevel growth model, as well as the results of the LRT significance tests. While the quadratic model always fit the data better than the linear model, the piecewise model did not improve fit in math in 4th to 6th grade.

How sensitive are estimates of the gains from extending the school year to the assumption of linear growth?

Results presented in Figure 3 indicate that estimates of how extending the school year by one month will affect learning can be highly sensitive to the assumption of linearity. While some grade-subject combinations show little difference between linear and quadratic estimates, others show stark differences. For example, looking at 2nd grade reading, linear estimates show that extending the school year by a month leads to a gain of 0.15 SD compared to 0.04 for the quadratic model. Further, whereas linear models indicate at least some gain in reading by extending the school year in 4th to 8th grades, the quadratic model suggests practically no gain.

How sensitive are estimates of summer learning loss to the assumption of linearity?

Figure 4 shows estimated summer loss by grade and subject based on linear versus quadratic extrapolations. Results are reported in terms of the SD of the spring test score. As the figure illustrates, for several grade-subject combinations, the discrepancies between results assuming different within-year functional forms are minimal. For instance, in 5th grade mathematics (the largest estimated summer loss by subject-grade combination), linear and quadratic estimates are virtually identical. However, for other grade-subject pairings, differences in results appear practically significant, especially for reading in the early grades. As an example, in reading for grades one and two, the linear estimates suggest that summer loss exceeds 0.10 SD. Yet, based on the quadratic model, the summer loss estimates for those grades are roughly half as large. Thus, while not all summer loss estimates appear sensitive to the assumption of within-year linearity, some certainly do.

To the degree we see nonlinear growth, are certain racial/ethnic and gender groups more likely to show a relative slowdown in learning rates across the school year?

Table 6 presents results from our multilevel model with controls for race/ethnicity, gender, and school percentage FRPL. When examining the quadratic parameter estimate and its association with race, results are highly dependent on the grade in question. In fact, the results indicate that Black and Hispanic students show less within-year deceleration in early grades, but more deceleration in later grades. For example, in kindergarten, the quadratic fixed effect estimate is -0.07 and the coefficient on Hispanic is 0.03 (significant at the .05 level). Thus, the deceleration is less pronounced for Hispanic kindergarten students relative to White students. By comparison, in 8th grade Hispanic students' growth is decelerating at nearly double the rate of White students, though overall deceleration rates are low for both groups. However, one should note that this finding did not hold across all racial/ethnic groups. Unlike for Black and Hispanic students, the deceleration for Asian students is almost always less extreme relative to White students, regardless of grade.

Within-year deceleration is also associated with gender in our sample. In most grades, deceleration is larger for males, especially in later grades. For example, in 8th grade, the male students demonstrate twice the rate of deceleration as female students. By contrast, the differences between male and female students are either nonsignificant or small for kindergarten and 1st grade.

Discussion

Many inferences related to program evaluation and policy require an accurate estimate of how much learning occurs during the school year. Yet, because most achievement tests are administered once a year in the spring, and few assessments are administered more than two timers per year, most related research simply assumes that within-year growth is linear. While a handful of studies provide some evidence germane to the tenability of this assumption in kindergarten and 1st grade (Fitzpatrick, Grissmer, & Hastedt, 2011; Hayes & Gershenson, 2016), we are aware of no studies that thoroughly investigate the assumption of the within-year linearity across upper elementary and middle school grades. Given that preparation for federally-mandated standardized testing in the spring may drive the sequencing of instruction in these later grades, it is highly plausible that learning rates prior to and subsequent to spring end-of-grade testing are different.

In this study, we address this gap in the literature by using a test typically administered three times per year to over seven million students nationwide. Using those data, we are able to estimate models that assume within-year linearity, as well as models that relax the assumption through inclusion of a within-year quadratic term, and compare results. Specifically, results are compared in terms of not only model fit, but also the sensitivity of policy-relevant estimates like the effect of extending the school year, summer loss, and the development of achievement gaps.

These analyses provide results that should be helpful not only to methodologists working with K-8 assessment data, but also applied researchers interested in related questions of practice, policy, and evaluation.

First, our results suggest that the assumption of linear within-year growth is tenuous at best, with the most extreme violations occurring for reading gains within late elementary and middle school grades. While the linear growth assumption appears plausible in most grades for math, growth appears to decelerate during the school year in reading across most grades. The differences between the fall-winter and winter-spring effect sizes in math are as high as 0.17 SD in kindergarten. In middle school, fall-winter growth effect sizes in reading are approximately twice as large as the effect sizes observed between winter and spring. These results are further borne out when multilevel models are fit to the data, with significant quadratic terms in all grade-year combinations.

Second, the questionable assumption of within-year linearity has implications for our understanding of how extending the school year might affect learning. That is, the estimated effect of extending the school year by one month is highly sensitive to the assumption of linearity. For instance, in 2^{nd} grade reading, linear estimates suggest extending the school year leads to a gain of 0.08 SD compared to 0.03 SD for the quadratic model. Similarly, while linear models indicate benefits to extending the school year in 2^{nd} through 8^{th} grade for learning reading, the quadratic model suggests gains are very near zero. Given much of the current justification for extending the school year assumes students will continue to learn at the same rate during the extra time added, our findings indicate that a portion of the related research base relies on an assumption that is often not justifiable.

Third, we see that summer loss estimates can be sensitive to the within-year linearity assumption. While some estimates of summer loss in math change little between linear and quadratic extrapolations, fairly large shifts occur in reading. As an example, for reading during the early grades, quadratic model estimates of summer loss are about half those same estimates when based on the linear model. Given these results, the common practice of estimating the learning that takes place between a spring achievement test administration and the end of the school year using linear extrapolation often may lead to overestimates of summer learning loss. Further, results could be even more sensitive to the linearity assumption if the learning that occurs after testing finishes slows even more precipitously than prior to testing in ways that cannot be captured with our data (i.e., we do not have test scores on the last day of the school year that would enable such inferences).

Finally, we show that the rate of deceleration is associated with race and gender after controlling for school-level socioeconomic status. For Black and Hispanic students, deceleration is less pronounced in early grades but more pronounced in later grades. This finding has implications for understanding achievement gaps and how they develop as students move through school. Specifically, though not entirely consistent by grade and race, our results indicate that gaps increase somewhat towards the end of the school year during middle school, when growth slows more for racial minority students than for White students. We also find that, for virtually every grade, growth decelerates more for males than females. While our results are in no way causal, they are commensurate with a theory that the progression of achievement gaps may be slowed by reducing declines in learning that occur at the end of the school year for male and racial minority students.

In sum, we show that assuming linear within-year growth may be a threat to many inferences researchers want to make about how certain programs and policies shape students' reading development. Given these findings, there are at least a couple of options for applied researchers interested in such inferences. First, studies that involve questions about how much learning occurs during the school year would ideally use scores from tests administered three or more times per year. While such tests are not currently the norm, they are expanding due to the growing prevalence of commercial benchmark assessments like MAP Growth and consortia tests like the Smarter Balanced Assessment Consortium (SBAC) assessments, which include an interim assessment option. Second, if more than two test scores per year are not available, then researchers should at minimum be aware that their estimates of summer loss and the benefits of extending the school year are likely overstatements for certain subject-grade combinations. Thus, researchers may wish to indicate that results assuming linear growth represent upper bounds on effects related to summer loss and changing the length of the school year.

Limitations

There are a few limitations to our study that bear mention. For one, our results are based on only a single assessment. While MAP Growth is currently utilized by roughly one in four public schools serving kindergarten to 8th grade, one cannot be certain results would generalize to other tests. Thus, replication of our findings with other measures is warranted (while noting that part of the reason there is so little evidence on the within-year linearity assumption is that few such tests/datasets exist).

Another limitation is that, while we do have three tests per year for most students, we do not have tests offered at the very beginning or end of the school year. That is, one might ideally have five tests: the three that we base our results on, and tests administered right when students arrive at school in the fall and just before they leave for summer. However, this limitation of our data could very well mean that our findings are understated, not overstated. For example, if schools greatly reduce instruction immediately following spring testing, then deceleration in learning would be even greater, and occur in ways we cannot examine with our data.

Finally, school districts set their own testing calendars due to potentially idiosyncratic reasons. Further research should investigate whether observed student and school characteristics explain the large variability observed in testing dates (and correspondingly months of instruction). Such research could also consider whether these differences impact our understanding of students' within-year growth trajectories on MAP Growth and other assessments.

Conclusion

A wide range of studies provide evidence to inform policy and practice decisions like whether to extend the school year and how best to reduce achievement gaps. Many such studies rely on estimates of how much learning occurs during the school year, and typically assume that learning occurs at a constant rate throughout the school year. However, our findings indicate that this assumption is often not justifiable, particularly in reading. In some cases, fall to winter gains are much larger than winter to spring gains. Further, we show that estimates of summer learning loss, changes in achievement gaps, and the effect of extending the school year can be quite sensitive to the assumption of within-year linearity. Thus, much of what we know on related topics may rely on a faulty assumption about how learning occurs during the year.

References

Atteberry, A., & McEachin, A. (2016). School's Out: The Role of Summers in Understanding Achievement Disparities. (EdWorkingPaper: 19-82). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/2mam-bp02

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 289–328.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). The academic achievement gap in grades three to eight. *NBER Working Paper, 12207*.

Dadey, N. & Briggs, D. (2012). A meta-analysis of growth trends from vertically scaled assessments. *Practical Assessment, Research, and Evaluation, 17* (14), 1-12.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*, 613–635.

Ferguson, R. (2001). Test score trends along racial lines, 1971-1996: Popular culture and community academic standards. *America Becoming: Racial Trends and Their Consequences*, *1*, 348–90.

Fitzpatrick, M.D., Grissmer, D., Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, *30* (2), 269–279.

Hanushek, E. A., & Rivkin, S. G. (2006). *School quality and the black-white achievement gap*. National Bureau of Economic Research. http://www.nber.org/papers/w12651

Hayes, M.S. & Gershenson, S. (2016). What differences a day can make: Quantile regression estimates of the distribution of daily learning gains. *Economics Letters*, *141*, 48–51.

Husain, M., & Millimet, D. L. (2009). The mythical 'boy crisis'? *Economics of Education Review*, *28*(1), 38–48.

LoGerfo, L., Nichols, A., & Chaplin, D. (2006). *Gender gaps in math and reading gains during elementary and high school by race and ethnicity*. Washington, DC: Urban Institute.

Murnane, R. J., Willett, J. B., Bub, K. L., McCartney, K., Hanushek, E., & Maynard, R. (2006). Understanding trends in the black-white achievement gaps during the first years of school [with comments]. *Brookings-Wharton Papers on Urban Affairs*, 97–135.

Neal, D. (2006). Why has black–white skill convergence stopped? *Handbook of the Economics of Education*, *1*(1), 511–576.

NWEA. (2020). *Case Studies*. Retrieved from https://www.nwea.org/case-studies/.

Patall, E. A., Cooper, H. Allen, A. B. (2010). Extending the school day or school year: A systematic review of research (1985 – 2009), *Review of Educational Research, 80*(3), 401 – 436.

Phillips, M., Crouse, J., & Ralph, J. (1998). Does the Black-White test score gap widen after children enter school. *The Black-White Test Score Gap*, 229–272.

Quinn, D. M. (2015). Black–white summer learning gaps: Interpreting the variability of estimates across representations. *Educational Evaluation and Policy Analysis, 37*(1), 50–69.

Quinn, D. M., Cooc, N., McIntyre, J., & Gomez, C. J. (2016). Seasonal dynamics of academic achievement inequality by socioeconomic status and race/ethnicity: Updating and extending past research with new national data. *Educational Researcher*, *45*, 443–453.

Rathbun, A., West, J., & Hausken, E. G. (2004). From Kindergarten through third grade children's beginning school experiences. Washington, DC: National Center for Education Statistics.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks: Sage.

Raudenbush, S.W., Bryk, A.S, Cheong, Y.F. & Congdon, R. (2011). *HLM 7 for Windows* [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853–891.

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*(2), 268–302.

Rocha, E. (2008). *Expanded Learning Time in Action: Initiatives in High-Poverty and High-Minority Schools and Districts*. Washington DC: Center for American Progress and Institute for America's Future. Retrieved from https://www.americanprogress.org/wp-content/uploads/issues/2008/07/pdf/elt1.pdf

von Hippel, P. T., & Hamrock, C. (2019). Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them. *Sociological Science, 6*, 43–80.

von Hippel, P. T., Workman, J., & Downey, D. B. (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*, 323–357.

Table 1
Descriptive Statistics for the Analytic Sample

		Race/ethnicity					
Grade	N	White	Black	Asian	Hispanic	Other race	Male
K	582,680	0.45	0.18	0.04	0.17	0.17	0.51
1	687,784	0.46	0.18	0.04	0.17	0.16	0.51
2	845,006	0.46	0.18	0.04	0.17	0.15	0.51
3	911,229	0.46	0.17	0.04	0.18	0.15	0.51
4	900,093	0.47	0.17	0.04	0.17	0.14	0.51
5	916,657	0.48	0.17	0.04	0.18	0.14	0.51
6	884,316	0.47	0.17	0.04	0.17	0.14	0.51
7	858,296	0.48	0.16	0.04	0.18	0.14	0.51
8	842,298	0.48	0.16	0.04	0.18	0.14	0.51

Note. N=the number of unique students within each grade in 2017-18.

Table 2
Comparison of the NWEA Sample of Schools and the Population of US Public Schools Serving Kindergartners through 8th Graders

	NWEA Schools			Population of Schools Serving K-8		
	N	M	SD	N	M	SD
Kindergarten	11,832	68.30	41.66	43,189	66.52	45.72
1st grade	11,960	69.19	40.94	43,533	67.23	44.12
2nd grade	12,016	69.63	40.92	43,601	67.49	44.07
3rd grade	12,042	72.21	43.12	43,599	69.83	46.37
4th grade	11,951	73.71	46.25	43,364	71.22	48.90
5th grade	11,552	76.66	55.92	42,142	73.54	58.40
6th grade	7,674	105.64	108.96	30,568	99.51	113.28
7th grade	6,394	123.10	122.23	27,323	109.86	125.60
8th grade	6,305	124.44	124.44	27,372	109.58	126.78
Percent FRPL	16,824	0.49	0.30	62,576	0.50	0.31
Percent Hispanic	16,824	0.17	0.22	62,597	0.20	0.24
Percent Black	16,824	0.18	0.26	62,597	0.16	0.25
Percent White	16,824	0.55	0.33	62,597	0.55	0.33
Percent Asian	16,824	0.04	0.07	62,597	0.03	0.07
City	16,824	0.27	0.45	62,601	0.25	0.43
Suburb	16,824	0.34	0.47	62,601	0.32	0.47
Town	16,824	0.12	0.33	62,601	0.13	0.34
Rural	16,824	0.27	0.44	62,601	0.30	0.46

Note. N=count of the number of schools with observed data for a given variable, M=mean, SD=standard deviation, FRPL=free or reduced priced lunch. The enrollment variables (kindergarten through 8th grade) report the number of students enrolled in each grade, and the associated counts represent the number of schools enrolling students in that grade within the sample and population. All school characteristics are from the 2017-18 Common Core of Data from the National Center of Education Statistics.

Table 3

Descriptive Statistics for Students' Months of Exposure and MAP Growth Test Scores by Subject, Grade, and Term

		Math					Reading				
			Months of School		RIT			Months of School		RIT	
Grade	Term	N	M	SD	M	SD	N	M	SD	M	SD
K	F17	473,238	0.95	0.50	141.13	10.44	445,460	0.87	0.50	138.41	9.71
	W18	516,842	4.66	0.75	151.26	12.15	499,713	4.60	0.77	147.43	11.33
	S18	575,289	8.52	0.60	160.20	12.70	550,741	8.46	0.63	155.48	12.67
1	F17	666,922	0.86	0.50	160.11	12.82	626,080	0.77	0.47	156.27	13.11
	W18	618,445	4.63	0.75	169.74	13.01	593,025	4.55	0.76	165.46	14.01
	S18	691,573	8.53	0.62	177.45	13.73	655,040	8.45	0.64	172.53	14.28
2	F17	800,435	0.82	0.49	175.69	13.78	772,175	0.74	0.46	172.88	16.25
	W18	736,288	4.60	0.77	183.42	13.42	728,549	4.52	0.77	180.89	16.33
	S18	817,522	8.52	0.64	190.13	13.97	796,421	8.43	0.68	186.44	16.07
3	F17	845,222	0.77	0.50	188.77	13.64	842,241	0.71	0.48	187.37	17.02
	W18	769,286	4.55	0.77	195.70	13.48	787,117	4.48	0.76	193.72	16.42
	S18	801,432	8.46	0.70	201.79	14.35	811,509	8.39	0.72	197.72	16.38
4	F17	831,191	0.74	0.49	200.63	14.27	828,996	0.68	0.46	197.53	16.69
	W18	763,519	4.54	0.77	205.85	14.34	773,219	4.47	0.75	202.22	15.99
	S18	788,121	8.47	0.72	211.80	15.82	794,897	8.39	0.75	205.17	16.08
5	F17	843,864	0.73	0.51	210.13	15.69	839,039	0.68	0.49	204.92	16.47
	W18	770,486	4.53	0.78	214.76	16.12	776,493	4.46	0.75	208.77	15.73
	S18	791,055	8.45	0.72	220.03	17.65	794,579	8.38	0.75	211.02	15.82
6	F17	814,015	0.73	0.49	214.87	16.14	809,928	0.70	0.47	210.30	16.39
	W18	698,556	4.53	0.78	218.50	16.69	705,254	4.49	0.77	212.91	15.97
	S18	746,959	8.41	0.74	222.82	17.73	747,312	8.34	0.76	214.93	16.01
7	F17	772,057	0.75	0.51	221.47	17.76	773,826	0.72	0.49	214.65	16.48
	W18	642,492	4.54	0.80	224.10	18.21	648,762	4.50	0.78	216.49	16.27
	S18	704,094	8.41	0.75	228.10	19.01	702,326	8.33	0.78	218.47	16.30
8	F17	739,820	0.75	0.51	226.90	18.91	752,177	0.72	0.48	218.44	16.54
	W18	620,204	4.52	0.79	229.09	19.15	638,877	4.49	0.79	220.17	16.22
	S18	660,222	8.39	0.75	232.53	20.21	666,113	8.30	0.77	221.69	16.33

Note. F17=fall of 2017, W18=winter of 2018, S18=spring of 2018, N=count of the number of schools with observed data for a given variable, M=mean, SD=standard deviation.

Table 4

Effect Sizes and Average Monthly Learning Gains by Subject and Grade

Subject	Grade	N	Months Between Test Dates		Growth Effect Sizes		Average Monthly RIT Gains	
			Fall-Winter	Winter-Spring	Fall-Winter	Winter-Spring	Fall-Winter	Winter-Spring
Math	K	380,431	3.69	3.90	0.93	0.76	3.18	2.74
	1	535,379	3.79	3.91	0.78	0.59	2.88	2.33
	2	619,387	3.80	3.95	0.61	0.49	2.54	2.13
	3	614,560	3.77	3.98	0.56	0.44	2.35	1.99
	4	602,242	3.78	4.00	0.40	0.39	2.02	2.00
	5	597,556	3.79	4.00	0.33	0.31	1.96	1.97
	6	534,229	3.77	4.00	0.27	0.24	1.87	1.78
	7	466,753	3.77	3.98	0.20	0.18	1.80	1.71
	8	438,221	3.76	3.98	0.17	0.13	1.79	1.68
Reading	K	361,805	3.70	3.91	0.90	0.72	3.03	2.69
	1	514,473	3.80	3.91	0.71	0.51	2.91	2.34
	2	610,337	3.81	3.92	0.53	0.34	2.83	2.16
	3	627,460	3.77	3.96	0.41	0.24	2.60	2.02
	4	608,889	3.78	3.98	0.32	0.18	2.29	1.85
	5	601,848	3.78	3.98	0.27	0.14	2.13	1.74
	6	531,326	3.77	3.95	0.20	0.11	2.01	1.75
	7	467,379	3.76	3.93	0.16	0.08	1.96	1.76
	8	445,427	3.75	3.92	0.15	0.06	1.95	1.73

Note. The months between tests is reported as the average number of months between a student's fall and winter (or winter and spring) tests. Growth effect sizes and the projected additional gains are reported in standard deviation units.

Table 5

Parameter Estimates from the Unconditional Multilevel Growth Models

Grade	Parameter	Math			Reading		
		Linear Model	Quadratic Model	Piecewise Model	Linear Model	Quadratic Model	Piecewise Model
K	Intercept	139.09 (0.06)	138.15 (0.06)	127.53 (0.07)	136.78 (0.05)	136.11 (0.05)	126.46 (0.06)
	Linear	2.52 (0.01)	3.13 (0.02)		2.25 (0.01)	2.71 (0.02)	
	Quadratic		-0.06 (0.00)			-0.05 (0.00)	
	Linear – Fall			2.81 (0.01)			2.48 (0.01)
	Linear – Spring			2.28 (0.01)			2.06 (0.01)
1	Intercept	158.49 (0.06)	157.45 (0.06)	148.86 (0.07)	155.00 (0.07)	154.02 (0.06)	146.13 (0.07)
	Linear	2.26 (0.00)	3.07 (0.01)		2.12 (0.00)	2.92 (0.01)	
	Quadratic		-0.09 (0.00)			-0.09 (0.00)	
	Linear – Fall			2.65 (0.01)			2.50 (0.01)
	Linear – Spring			1.87 (0.01)			1.72 (0.01)
2	Intercept	174.16 (0.06)	173.46 (0.06)	166.11 (0.07)	171.84 (0.07)	170.57 (0.07)	165.06 (0.08)
	Linear	1.89 (0.00)	2.47 (0.01)		1.78 (0.00)	2.84 (0.01)	
	Quadratic		-0.06 (0.00)			-0.12 (0.00)	
	Linear – Fall			2.17 (0.01)			2.28 (0.01)
	Linear – Spring			1.59 (0.01)			1.23 (0.01)
3	Intercept	187.39 (0.06)	186.74 (0.06)	180.12 (0.06)	186.54 (0.07)	185.42 (0.07)	181.57 (0.08)
	Linear	1.71 (0.00)	2.23 (0.01)		1.39 (0.00)	2.34 (0.01)	
	Quadratic		-0.06 (0.00)			-0.10 (0.00)	
	Linear – Fall			1.95 (0.01)			1.84 (0.01)
	Linear – Spring			1.43 (0.01)			0.87 (0.01)
4	Intercept	199.11 (0.07)	199.09 (0.07)	192.32 (0.07)	196.87 (0.07)	196.00 (0.07)	193.32 (0.08)
	Linear	1.45 (0.00)	1.47 (0.01)		1.02 (0.00)	1.77 (0.01)	
	Quadratic		0.00 (0.00)			-0.08 (0.00)	
	Linear – Fall			1.46 (0.01)			1.38 (0.01)
	Linear – Spring			1.42 (0.01)			0.61 (0.01)
5	Intercept	208.53 (0.08)	208.46 (0.08)	202.52 (0.07)	204.19 (0.07)	203.39 (0.08)	201.46 (0.08)
	Linear	1.29 (0.00)	1.35 (0.01)		0.82 (0.00)	1.51 (0.01)	
	Quadratic		-0.01 (0.00)			-0.08 (0.00)	
	Linear – Fall			1.32 (0.01)			1.15 (0.01)
	Linear – Spring			1.25 (0.01)			0.45 (0.01)
6	Intercept	212.72 (0.10)	212.56 (0.10)	207.98 (0.10)	208.92 (0.10)	208.36 (0.10)	206.69 (0.10)
	Linear	1.06 (0.00)	1.20 (0.01)		0.65 (0.00)	1.13 (0.01)	
	Quadratic		-0.02 (0.00)			-0.05 (0.00)	
	Linear – Fall			1.14 (0.01)			0.88 (0.01)
	Linear – Spring			0.97 (0.01)			0.38 (0.01)
7	Intercept	219.10 (0.12)	219.06 (0.12)	215.23 (0.12)	213.22 (0.10)	212.77 (0.10)	211.32 (0.11)
	Linear	0.86 (0.01)	0.91 (0.01)		0.53 (0.00)	0.91 (0.02)	
	Quadratic		-0.01 (0.00)			-0.04 (0.00)	
	Linear – Fall			0.89 (0.01)			0.69 (0.01)
	Linear – Spring			0.81 (0.01)			0.34 (0.01)
8	Intercept	224.69 (0.13)	224.42 (0.13)	221.64 (0.13)	217.03 (0.10)	216.57 (0.11)	215.63 (0.11)
	Linear	0.73 (0.01)	0.96 (0.02)		0.46 (0.01)	0.86 (0.02)	
	Quadratic		-0.03 (0.00)			-0.05 (0.00)	
	Linear – Fall			0.82 (0.01)			0.65 (0.01)
	Linear – Spring			0.61 (0.01)			0.22 (0.01)

Note. All parameters reported in this table are statistically significant ($p<.001$).

Table 6(a)

Results from the Conditional Growth Model Predicting Math Score Deceleration Based on Student and School Characteristics

Parameter	K	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade
Intercept	140.272 (0.054)	159.841 (0.058)	175.485 (0.058)	188.619 (0.056)	200.962 (0.060)	210.726 (0.068)	215.272 (0.093)	222.600 (0.112)	228.483 (0.124)
Percent FRPL	-6.383 (0.172)	-7.534 (0.186)	-7.526 (0.188)	-8.578 (0.193)	-8.683 (0.212)	-9.258 (0.237)	-9.488 (0.316)	-9.009 (0.383)	-9.762 (0.425)
Black	-3.776 (0.067)	-5.622 (0.073)	-5.860 (0.077)	-6.351 (0.072)	-6.961 (0.077)	-8.086 (0.086)	-8.630 (0.117)	-9.617 (0.130)	-10.335 (0.149)
Asian	-1.044 (0.136)	*-0.011 (0.139)*	2.647 (0.121)	2.968 (0.120)	3.832 (0.129)	4.466 (0.144)	4.626 (0.200)	5.517 (0.247)	5.837 (0.266)
Hispanic	-4.700 (0.073)	-5.518 (0.077)	-5.015 (0.077)	-4.910 (0.076)	-5.105 (0.081)	-5.808 (0.086)	-6.472 (0.122)	-7.382 (0.142)	-8.271 (0.163)
Other Race	-1.922 (0.075)	-2.389 (0.079)	-2.061 (0.078)	-2.163 (0.079)	-2.440 (0.081)	-3.068 (0.089)	-3.100 (0.111)	-3.775 (0.144)	-4.091 (0.154)
Male	-0.742 (0.031)	-0.316 (0.034)	*0.014 (0.034)*	0.437 (0.033)	0.694 (0.034)	0.758 (0.036)	0.414 (0.042)	-0.320 (0.047)	-0.756 (0.050)
Linear Slope	3.236 (0.019)	3.024 (0.014)	2.416 (0.014)	2.171 (0.014)	1.381 (0.012)	1.265 (0.012)	1.178 (0.014)	0.881 (0.015)	0.897 (0.018)
Percent FRPL	-0.222 (0.055)	-0.071 (0.041)	-0.135 (0.040)	0.191 (0.036)	*0.056 (0.035)*	0.074 (0.035)	-0.117 (0.044)	-0.240 (0.049)	0.053 (0.056)
Black	-0.459 (0.024)	-0.183 (0.019)	-0.145 (0.018)	*-0.027 (0.021)*	0.027 (0.016)	-0.030 (0.017)	-0.127 (0.018)	-0.053 (0.019)	-0.052 (0.028)
Asian	-0.003 (0.041)	-0.042 (0.032)	-0.203 (0.027)	-0.163 (0.027)	*-0.011 (0.024)*	0.114 (0.025)	0.060 (0.025)	*0.035 (0.030)*	*-0.016 (0.034)*
Hispanic	-0.360 (0.025)	-0.083 (0.019)	-0.114 (0.017)	*0.003 (0.021)*	0.028 (0.016)	0.049 (0.016)	-0.065 (0.017)	-0.033 (0.019)	0.071 (0.029)
Other Race	-0.141 (0.026)	-0.037 (0.021)	-0.072 (0.019)	*-0.011 (0.022)*	*-0.009 (0.017)*	*0.016 (0.018)*	*-0.013 (0.019)*	-0.036 (0.022)	*0.001 (0.031)*
Male	0.087 (0.013)	0.201 (0.010)	0.231 (0.009)	0.151 (0.009)	0.161 (0.008)	0.151 (0.009)	0.126 (0.010)	0.125 (0.010)	0.127 (0.012)
Quadratic Slope	-0.071 (0.002)	-0.086 (0.001)	-0.059 (0.001)	-0.050 (0.001)	0.008 (0.001)	0.009 (0.001)	-0.005 (0.002)	0.003 (0.002)	-0.014 (0.002)
Percent FRPL	0.022 (0.005)	*-0.005 (0.004)*	0.011 (0.004)	-0.018 (0.004)	-0.019 (0.004)	-0.024 (0.004)	*0.005 (0.005)*	0.019 (0.005)	*-0.003 (0.006)*
Black	0.026 (0.002)	0.005 (0.002)	0.002 (0.002)	-0.010 (0.002)	-0.020 (0.002)	-0.016 (0.002)	*-0.001 (0.002)*	-0.004 (0.002)	-0.005 (0.003)
Asian	*0.004 (0.004)*	0.016 (0.003)	0.020 (0.003)	0.024 (0.003)	0.013 (0.003)	0.005 (0.003)	0.010 (0.003)	0.013 (0.003)	0.018 (0.004)
Hispanic	0.028 (0.002)	0.005 (0.002)	0.008 (0.002)	*-0.002 (0.002)*	-0.010 (0.002)	-0.014 (0.002)	*0.000 (0.002)*	*-0.001 (0.002)*	-0.011 (0.003)
Other Race	0.007 (0.003)	*0.000 (0.002)*	*0.003 (0.002)*	-0.005 (0.002)	-0.005 (0.002)	-0.009 (0.002)	-0.004 (0.002)	*-0.001 (0.002)*	*-0.005 (0.003)*
Male	*0.002 (0.001)*	-0.005 (0.001)	-0.014 (0.001)	-0.010 (0.001)	-0.012 (0.001)	-0.018 (0.001)	-0.021 (0.001)	-0.020 (0.001)	-0.017 (0.001)

Note. Gray text represents parameters that are not statistically significant at a .05 level.

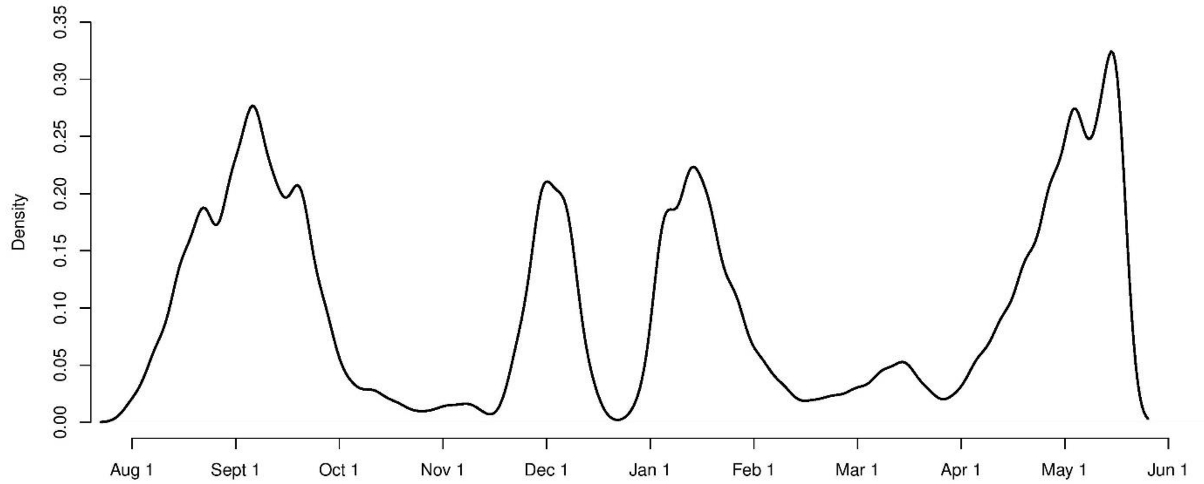
Table 6(b)

Results from the Conditional Growth Model Predicting Reading Score Deceleration Based on Student and School Characteristics

Parameter	K	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade
Intercept	138.117 (0.053)	156.992 (0.062)	173.883 (0.070)	189.225 (0.069)	199.934 (0.069)	207.503 (0.069)	212.718 (0.091)	217.461 (0.101)	221.510 (0.107)
Percent FRPL	-4.851 (0.164)	-7.172 (0.194)	-9.120 (0.222)	-10.131 (0.228)	-10.112 (0.231)	-9.826 (0.238)	-8.911 (0.313)	-7.992 (0.348)	-7.314 (0.368)
Black	-2.601 (0.065)	-4.467 (0.075)	-4.743 (0.090)	-5.863 (0.091)	-6.440 (0.094)	-6.885 (0.096)	-7.035 (0.125)	-7.135 (0.140)	-7.458 (0.138)
Asian	-0.963 (0.129)	-0.660 (0.153)	2.207 (0.144)	1.847 (0.135)	1.637 (0.135)	1.335 (0.133)	1.408 (0.167)	1.608 (0.177)	1.611 (0.183)
Hispanic	-3.742 (0.074)	-5.319 (0.081)	-5.718 (0.092)	-6.282 (0.094)	-6.443 (0.099)	-6.492 (0.094)	-6.736 (0.127)	-6.962 (0.137)	-7.236 (0.147)
Other Race	-1.410 (0.074)	-2.067 (0.082)	-1.794 (0.091)	-2.070 (0.095)	-2.491 (0.095)	-2.671 (0.095)	-2.660 (0.113)	-2.964 (0.136)	-2.991 (0.144)
Male	-1.501 (0.032)	-2.096 (0.035)	-2.795 (0.039)	-3.073 (0.039)	-3.001 (0.039)	-2.997 (0.039)	-3.272 (0.041)	-3.724 (0.043)	-4.026 (0.045)
Linear Slope	2.835 (0.020)	2.991 (0.015)	2.954 (0.015)	2.322 (0.014)	1.662 (0.013)	1.366 (0.012)	0.995 (0.016)	0.803 (0.018)	0.736 (0.018)
Percent FRPL	-0.325 (0.055)	-0.362 (0.041)	-0.387 (0.039)	-0.069 (0.037)	-0.033 (0.036)	0.114 (0.037)	-0.162 (0.051)	-0.099 (0.060)	-0.168 (0.059)
Black	-0.336 (0.025)	-0.233 (0.021)	-0.335 (0.020)	-0.123 (0.020)	0.025 (0.019)	0.085 (0.019)	0.084 (0.022)	0.034 (0.031)	0.112 (0.024)
Asian	-0.153 (0.045)	-0.014 (0.035)	-0.363 (0.031)	-0.314 (0.028)	-0.218 (0.027)	-0.094 (0.025)	0.031 (0.028)	-0.044 (0.032)	0.023 (0.031)
Hispanic	-0.282 (0.027)	-0.240 (0.022)	-0.239 (0.020)	-0.036 (0.020)	0.077 (0.019)	0.115 (0.019)	0.091 (0.022)	0.088 (0.032)	0.089 (0.025)
Other Race	-0.133 (0.027)	-0.105 (0.022)	-0.105 (0.022)	-0.080 (0.021)	0.003 (0.021)	0.024 (0.021)	0.035 (0.024)	0.025 (0.032)	0.008 (0.026)
Male	-0.006 (0.014)	0.038 (0.012)	0.016 (0.011)	0.145 (0.011)	0.199 (0.011)	0.220 (0.010)	0.207 (0.012)	0.170 (0.013)	0.197 (0.013)
Quadratic Slope	-0.052 (0.002)	-0.090 (0.001)	-0.126 (0.001)	-0.104 (0.001)	-0.074 (0.001)	-0.062 (0.001)	-0.039 (0.002)	-0.030 (0.002)	-0.032 (0.002)
Percent FRPL	0.016 (0.005)	0.021 (0.004)	0.037 (0.004)	0.013 (0.004)	0.011 (0.004)	-0.002 (0.004)	0.025 (0.005)	0.021 (0.006)	0.028 (0.006)
Black	0.013 (0.003)	0.011 (0.002)	0.023 (0.002)	0.006 (0.002)	-0.007 (0.002)	-0.010 (0.002)	-0.010 (0.002)	-0.004 (0.003)	-0.013 (0.003)
Asian	0.016 (0.004)	0.003 (0.004)	0.025 (0.003)	0.028 (0.003)	0.022 (0.003)	0.014 (0.003)	0.004 (0.003)	0.015 (0.003)	0.005 (0.003)
Hispanic	0.012 (0.003)	0.017 (0.002)	0.022 (0.002)	0.006 (0.002)	-0.004 (0.002)	-0.007 (0.002)	-0.005 (0.002)	-0.004 (0.003)	-0.004 (0.003)
Other Race	0.005 (0.003)	0.008 (0.002)	0.005 (0.002)	0.003 (0.002)	-0.003 (0.002)	-0.004 (0.002)	-0.005 (0.003)	-0.003 (0.003)	-0.003 (0.003)
Male	0.000 (0.001)	-0.004 (0.001)	0.002 (0.001)	-0.009 (0.001)	-0.016 (0.001)	-0.021 (0.001)	-0.024 (0.001)	-0.021 (0.001)	-0.023 (0.001)

Note. Gray text represents parameters that are not statistically significant at a .05 level.

(A) Distribution of Testing Dates



(B) Distribution of Months of Exposure to School

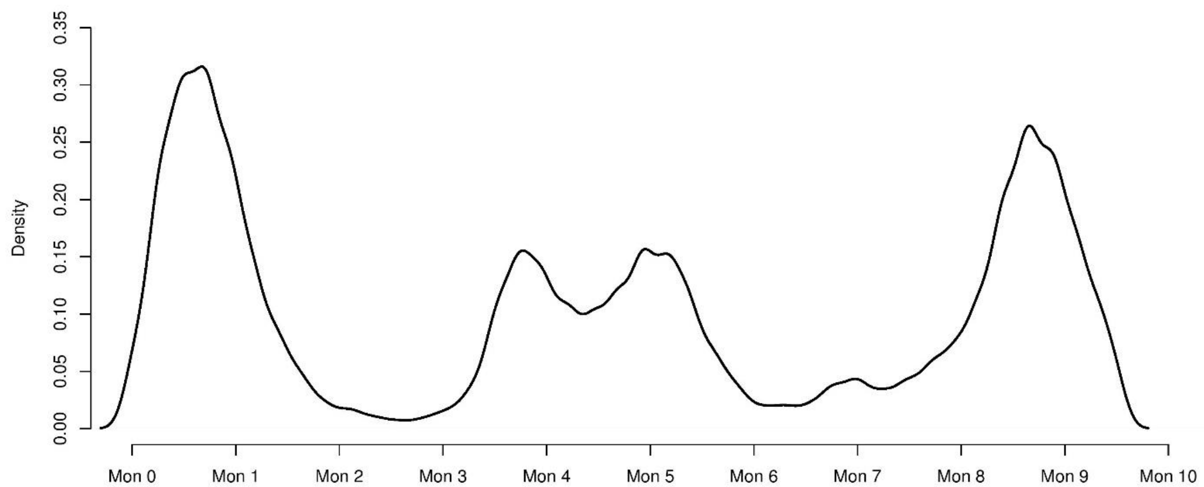


Figure 1. Distribution of test dates and months of exposure to school (e.g., months elapsed between the school start date and each student's testing date) during the 2017-18 school year. In this figure, we have pooled all students who tested in math at least once during the school year across grade levels.

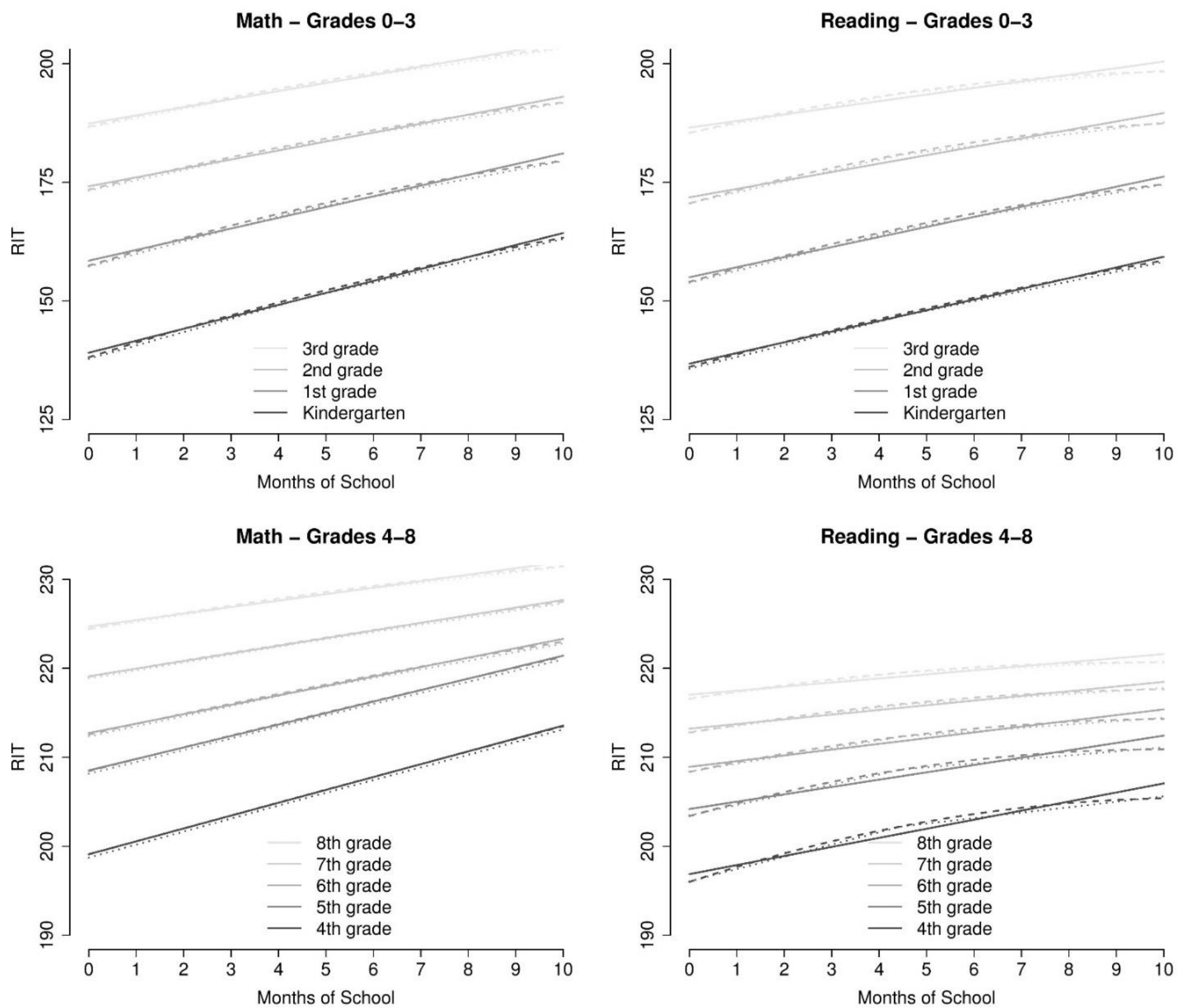


Figure 2. Estimated growth trajectories by grade, subject, and model type, where the solid line represents the linear growth model, the dashed line is the quadratic model, and the dotted line is the piecewise model. RIT is the metric of the MAP Growth assessments.

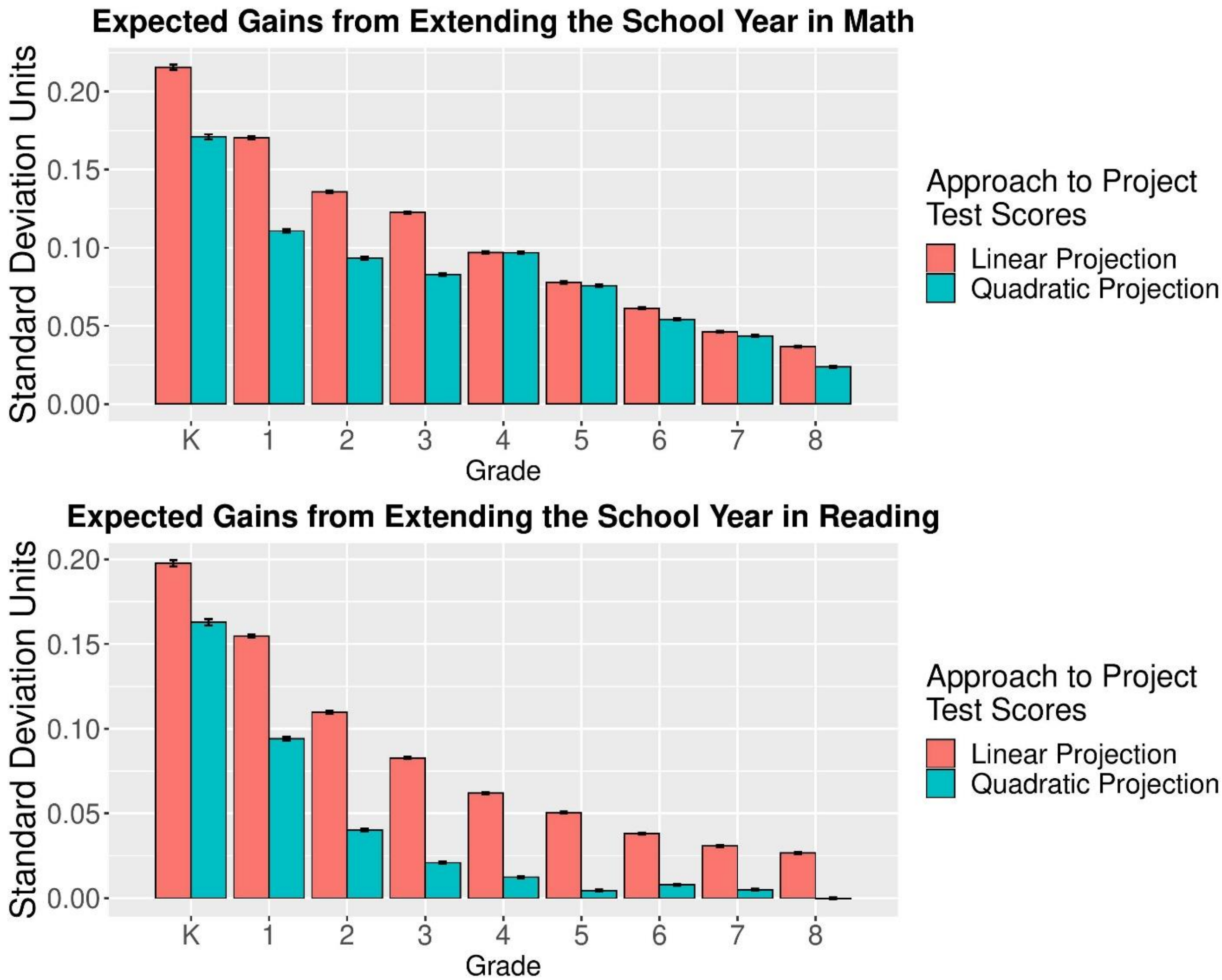


Figure 3. Expected gains from extending the school year by one month depending on whether growth is assumed to be linear or quadratic during the school year. The model-based additional gains estimates are the projected difference between students' average gains at nine or ten months of school exposure depending on whether student growth is modeled as a linear or quadratic function of time.   Note that we are not comparing schools that actually have nine versus ten-month school years; we are comparing model-based estimates of gains associated with going from a nine to a ten-month school calendar.

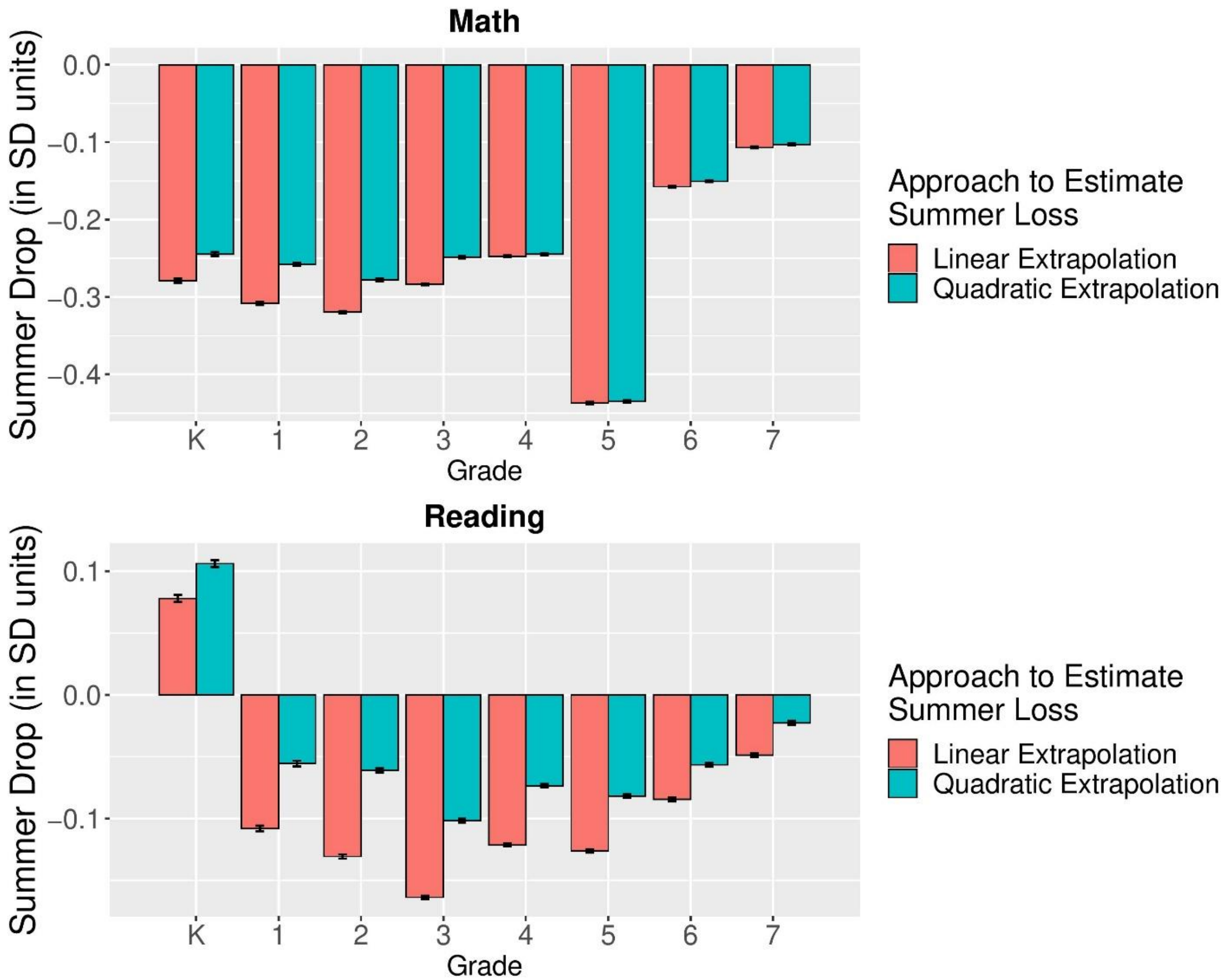


Figure 4. Estimated average summer learning loss by grade and subject under linear and quadratic test score projection. Results display the summer following each grade (e.g., K corresponds to the summer following kindergarten). Ninety-five percent confidence intervals are shown as black lines at the top/bottom of each bar. Summer loss estimates are reported in the unit of the spring 2018 standard deviation.

Supplemental Materials

Appendix A. Multilevel growth model fit comparison

Table A1

Sample Sizes and Model Fit Estimates for the Multilevel Growth Models

		Sample sizes			Deviance Estimates			Model Comparison Significance	
Subject	Grade	Level-1	Level-2	Level-3	Model 1 - Linear	Model 2 - Quadratic	Model 3 - Piecewise	Model 1 vs. 2	Model 1 vs. 3
Math	K	1,401,587	550,344	8,098	9929541.64	9912482.82	9908623.96	***	***
Math	1	1,747,753	652,742	9,507	12401200.08	12366245.97	12372883.08	***	***
Math	2	2,070,194	775,979	11,141	14726559.96	14689820.73	14694990.25	***	***
Math	3	2,134,882	810,726	11,414	15064166.53	15050627.11	15043689.70	***	***
Math	4	2,092,991	796,297	11,124	14824736.34	14806664.00	14826894.24	***	
Math	5	2,107,617	811,690	10,883	15196860.76	15182482.66	15206248.12	***	
Math	6	1,961,641	770,209	7,176	14204916.50	14193919.78	14206780.42	***	
Math	7	1,834,926	743,387	6,038	13538369.52	13529683.62	13537146.35	***	***
Math	8	1,754,261	720,415	5,967	13120900.13	13119370.11	13118169.21	***	***
Reading	K	1,336,591	521,697	7,835	9539869.54	9529108.18	9533990.18	***	***
Reading	1	1,653,637	609,917	9,072	11975847.98	11950700.11	11950775.78	***	***
Reading	2	2,020,153	746,052	10,973	15046502.68	15002087.80	15005958.93	***	***
Reading	3	2,160,991	808,115	11,456	16182670.03	16144692.70	16147461.41	***	***
Reading	4	2,105,126	792,252	11,115	15635832.38	15609990.67	15610007.92	***	***
Reading	5	2,112,212	805,117	10,881	15618657.34	15595638.20	15595549.41	***	***
Reading	6	1,966,038	768,734	7,173	14617404.99	14603371.68	14603688.49	***	***
Reading	7	1,841,210	743,356	5,984	13787809.15	13784343.48	13778632.64	***	***
Reading	8	1,785,931	729,023	5,929	13394487.53	13382494.90	13383111.64	***	***

Note. Deviance is equal to -2 times the estimated model log-likelihood.

*** p<0.001, ** p<0.01, * p<0.05

Appendix B. Calculation of Summer Learning Projections

For student *i* in school *j* in each subject/grade, the projected spring 2018 test score based on the linear model is ($\mathrm{RIT}^{PLin}{}_{Sij}$) is calculated as

$$\mathrm{RIT}^{PLin}{}_{Sij} = \mathrm{RIT}_{Sij} + \left(\hat{\gamma}_{100} + \hat{u}_{10j} + \hat{r}_{1ij}\right) * \mathrm{SpringMonths_{ij}},$$

where RIT_{Sij} is the observed spring test score student *i* in school j, $\hat{\gamma}_{100}$ is the estimated linear growth fixed effect from the multilevel growth model, $\hat{u}_{10j}$ is a school-level empirical Bayes (EB) estimate of the linear growth random effect, and $\hat{r}_{1ij}$ is student-level EB estimate of the linear growth random effect. $\mathrm{SpringMonths_{ij}}$ represents the months of school remaining in the 2017-18 school year when the student tested and is calculated by comparing the student's test date to the reported last day of school in the student's district. Under the quadratic model, the projected spring score is calculated as

$$\mathrm{RIT}^{PQua}{}_{Sij} = \mathrm{RIT}_{Sij} + \left(\hat{\gamma}_{100} + \hat{u}_{10j} + \hat{r}_{1ij}\right) * \mathrm{SpringMonths_{ij}} + \left(\hat{\gamma}_{200} + \hat{u}_{20j} + \hat{r}_{2ij}\right) * \mathrm{SpringMonths_{ij}^2},$$

where $\hat{\gamma}_{200}$ is the estimated quadratic growth fixed effect, $\hat{u}_{20j}$ is a school-level quadratic growth EB estimate, and $\hat{r}_{2ij}$ is student-level quadratic growth EB estimate.

Given that we did not observe students' trajectories across the 2018-19 school year, we did not have a similar model-based set of parameter estimates to project students' fall 2019 score. Given this limitation, we just make the naïve assumption that students' 2017-18 growth rate can be used as an approximation of expected growth during the first months of the subsequent school year. Therefore, the projected fall score ($\mathrm{RIT}^{P}{}_{Fij}$) that is used in the summer loss calculations for both models was estimated as

$$\mathrm{RIT}^{P}{}_{Fij} = \mathrm{RIT}_{Fij} + \left(\hat{\gamma}_{100} + \hat{u}_{10j} + \hat{r}_{1ij}\right) * \mathrm{FallMonths}_{\mathrm{ij}},$$

where RIT_{Fij} is the observed fall RIT score and $\mathrm{FallMonths}_{\mathrm{ij}}$ is the number of months that the student has been in school during the 2018-19 school prior to his or her fall test.

The projected summer loss for student *i* in school *j* under the linear growth projection is therefore

$$\Delta_{\mathrm{ij}}^{\mathrm{Lin}} = \mathrm{RIT}^{P}{}_{Fij} - \mathrm{RIT}^{PLin}{}_{Sij},$$

whereas the projected summer loss under the quadratic growth projection is

$$\Delta_{\mathrm{ij}}^{\mathrm{Qua}} = \mathrm{RIT}^{P}{}_{Fij} - \mathrm{RIT}^{PQua}{}_{Sij}.$$

Estimates for each projection approach are averaged within each grade/subject and then standardized by the spring 2018 standard deviation corresponding to the grade/subject combination.