ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-82

School's Out: The Role of Summers in Understanding Achievement Disparities

Allison Atteberry
University of Colorado Boulder

Andrew McEachin
RAND Corporation

The field is generally aware of the summer learning loss (SLL) phenomenon. However key characteristics of SLL are not broadly established. What proportion of students' school-year gains are lost in the subsequent summer? Is the magnitude of SLL generally similar across students or across grades? We describe the role summers play in the end-of-schooling achievement disparities using a unique dataset that spans eight grades, 200 million test scores, 18 million students, 50 states, and school-years 2008-2016. On average, 19% of students' pathways from their 1st to 8th grade test-score occur during summers. We show that—even if all inequality in school-year learning rates could be eliminated, students would still end school with very different achievement due to SLL alone.

VERSION: June 2019

Suggested citation: Atteberry, Allison, and Andrew McEachin. (2019). School's Out: The Role of Summers in Understanding Achievement Disparities. (EdWorkingPaper: 19-82). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-82

Introduction

Students exhibit differences in knowledge and preparation even in kindergarten (K) on the order of approximately 0.50 student test score standard deviations or six months of learning (Halle et al., 2009; Reardon, 2014; Shonkoff & Phillips, 2000). By the end of high school, these troubling disparities have only grown: Some students are ready to attend Ivy League colleges, while others struggle to pass exit exam tests of basic middle school math and reading skills. These disparate schooling endpoints launch students into adulthood with very different skillsets and opportunities and become a major contributor to the inequality in later life outcomes we see today in the US.

In this paper, we explore whether this "fanning out" of achievement occurs while students are in school or on summer break, using a novel dataset with over 200 million student test scores for students spread across the US. The answer to this question has important implications for where researchers and policy makers look for opportunities to disrupt this stratification process.

The field is generally aware of the phenomenon called summer learning loss (SLL)—that is, the fact that student learning slows during the summer. However key characteristics of SLL are not broadly established. For instance: Do students, on average, actually *lose* ground during the summer, or just exhibit no gain (i.e., flat)? What proportion of a student's school year gain tends to be lost in the summer that immediately follows? Is the magnitude of SLL generally similar across students? Or can some actually maintain the school-year learning rate throughout the summer (or conversely, do some lose all of their school-year gains in the shorter summer period)? Does this vary by grade level? Do the same students tend to experience SLL year after year, or are losses in one summer unrelated to losses in other summers? Ultimately, how large of a role does the summer period play in where a student ends up in the achievement distribution? This paper addresses these questions.

Even if, over the summer, students tend to lose ground, summers will only contribute to widening achievement disparities if students exhibit meaningful *variation* around the typical summer pattern. We therefore adopt a different approach from prior SLL in which we focus on estimating the degree of *variability* across students in SLL, relative to school-year learning gains.

Current Study

There have been logistical challenges to studying SLL: The annual end-of-school-year statewide testing systems often used by quantitative policy researchers simply lack fall datapoint needed to separate learning gains between the school year and the summer. Opportunities to investigate SLL have necessarily been limited to more idiosyncratic samples (e.g., one city), specific years, or particular grades (e.g., only after grade K). To illustrate this point, Figure 1 provides a summary of the years, sample sizes, grade levels, and geographic locations of data used in prior SLL research, alongside data for the current study.

Data provided by the Northwest Evaluation Association (NWEA) allow us to estimate means and variances in SLL across 8 grade levels, using over 200 million test scores for nearly 18 million students in 7,500 districts across all 50 states in a very recent time period (2008 through 2016). We use this powerful dataset in a hierarchical student growth modeling framework to characterize the contribution of SLL to end-of-school achievement disparities. Specifically, we answer the following four questions:

(1) *On average, how do average learning gains during the school year compare to gains/losses during the summer across grade levels?*
(2) *Of more relevance to the current investigation, how much do students vary in terms of how much they gain or lose?*
(3) *Do the same students tend to exhibit summer learning loss year after year, or are these gains/losses randomly distributed?*
(4) *What proportion of the variance in end-of-school outcomes arises during the summer?*

It should be noted that a different, major focus of SLL research has been to document its role in producing *racial* achievement gaps. Given the importance of that question, as well as and some of the methodological difficulties (see e.g., Quinn (2014) for a discussion), that crucial topic necessitates its own, separate, and full investigation. The current paper simply has a different goal:

To update the existing knowledge base about overall 1st through 8th grade school year learning gains and subsequent summer loss patterns, document the degree of variability in those patterns, and characterize the extent to which end-of-school achievement disparities arise during summers.

Students' test score trajectories follow a zig-zag pattern[1] of learning gains and losses between 1st and 8th grade that ultimately add up to how unequal outcomes are across students by the end of that period. This pathway between where a student starts and ends can be partitioned into two parts: gains made during the school years and gains/losses during the intervening summers. We find that, on average, 19 percent of a student's pathway from their 1st to 8th grade test score arises during the summers. For some students, this number is closer to 30 percent. Given that the summer is only about a quarter of a calendar year, we conclude that summers play an oversized role in 8th grade achievement inequality. With respect to the questions posed above, we *do* find that some students maintain their school-year learning rate throughout the summer, while others can lose almost as much ground as they had gained in the preceding school year. Ultimately, we show that—even if all the inequality in school-year learning rates could be entirely eliminated, students would still end school with very different achievement levels due to SLL alone.

In what follows, we first situate the contributions of the current study within existing SLL literature. Next, we introduce this remarkable dataset and how it compares to the broader U.S. public school population. We also describe a significant primary data collection activity undertaken to address a methodological concern in SLL research about the dates on which tests are taken (more on this below). In the Methods section, we present the specification of the multilevel model and the key parameters it estimates. The subsequent Results section is organized by the four research questions introduced above. The Conclusion provides a reflection on the findings, the study limitations, and implications for future research.

[1] Look ahead to Figure 2 for a hypothetical illustration of that zig-zag pattern.

Evidence on SLL

Children's Summer Time and Experiences

Research has documented that children experience vastly different home environments *prior* to formal schooling (Gilkerson & Richards, 2009; Kaushal, Magnuson, & Waldfogel, 2011; Kornrich & Furstenberg, 2013), and further that this pre-school time leads to sizeable achievement differences that are apparent on day one of kindergarten (Lee & Burkam, 2002; Magnuson, Meyers, Ruhm, & Waldfogel, 2004). It is often overlooked, however, that children spend much of their school-age years outside of school, as well. The majority of this time is concentrated in the summer months—a time when schools play little to no direct role in the organization of children's time use and activities. Instead, children return to full-time care of their families during the summer, and families have vastly different options and preferences for how children spend this time. In fact, the variety of environments and activities children spend their time on during the summer is likely much greater than during school (Gershenson, 2013). It is possible that student achievement gaps grow primarily during these summer months, when child experiences are probably most diverse.

Research on Summer Learning Loss

Much has been written about SLL (see e.g., Gershenson (2013) for a particularly thorough recent overview; or Cooper, Nye, Charlton, Lindsay, and Greathouse (1996) for a meta-analysis across early studies). Today, there is a common understanding among policy-makers, researchers, and practitioners that, during the summer, students lose some knowledge and skills acquired during the school year. The seminal research on summer setback comes from two key studies: Heyns' study of the summer after 6th grade for about 3,000 students in 42 Atlanta schools from 1970 to 1972 (Heyns, 1978), and Entwisle and Alexander's study of the summers after grades 1 and 2 for about 750 students in 20 Baltimore schools from 1982 to 1984 (Entwisle & Alexander, 1992). These studies documented the now-accepted conclusion that, on average, students tend to learn at slower rates during the summer. Heyns found that average 6th grade school year gains in Atlanta were positive (about 60 percent of a national norm for one year of achievement gains), while summer gains were either flat or very modestly negative, depending on cohort. Entwisle and Alexander (1992) used a multilevel, quadratic individual growth curve model to document slower summer (versus school-year) learning. The authors have continued to follow their Baltimore sample through adulthood and have found that early differences in summer learning are predictive of later life outcomes such as high school completion and college-going (Alexander, Entwisle, & Olson, 2007; Entwisle & Alexander, 1990; Entwisle & Alexander, 1992; Entwisle, Entwisle, & Olson, 2001).

The findings from these studies have become the definitive word on summer setback in the literature, raising awareness of the phenomenon and the role it plays in growing educational inequality. A series of studies have followed that examined SLL in specific locations (e.g., Allinder, Fuchs, Fuchs, & Hamlett (1992) in 2 rural schools around 1990; Borman, Benson, & Overman (2005) with about 300 students in Baltimore high poverty schools; Skibbe, Grimm, Bowles, & Morrison (2012) with about 380 students in 1 suburban Midwest town). That said, it has been unclear whether the results from those early studies would either generalize outside of their local contexts or to a vastly different educational landscape up to forty years later.

A handful of more recent studies have used the Early Childhood Longitudinal Study Kindergarten Class (ECLS-K) data from 1998-99 to at least move to a national sample in a more (Benson & Borman, 2010; Burkam, Ready, Lee, & LoGerfo, 2004; Downey, von Hippel, & Broh, 2004; Downey, von Hippel, & Hughes, 2008; Quinn, 2014). However, this dataset only covers one summer between K and 1st grade nearly two decades ago for one-third of the ECLS-K sample. Unfortunately, the ECLS-K data set does not provide an opportunity to observe if SLL increases or decreases as students progress through school.

In short, the existing SLL research base has been constrained by the availability of data. See Figure 1 for a summary of the data used across 12 key SLL studies. Figure 1 highlights how unique the current dataset is, relative to the limitations imposed by prior data. For instance, none of the data listed in Figure 1 are from the last decade—a time period with a significant amount of experimentation with teacher and school policies (e.g., high stakes accountability)—nor do they cover students in later elementary and middle school grades. Most datasets used previously have several hundred to a few thousand students (the largest dataset includes about 23 thousand students), compared to the over 17 million students for whom we observe test scores.

In sum, the extant research on summer learning loss took an important leap forward in the late twentieth century, and it now seems to be experiencing a resurgence of interest, particularly spurred by the availability of the ECLS-K data. This new work has sought to build upon the methods used in prior work (e.g., taking into account test timing), update the evidence to the late 1990s, and to cover a nationally representative sample (at least of kindergarteners).

The current paper continues in this tradition, building off the various methodological advances put forth in this domain. We also seek to contribute in a few additional, unique ways. First, NWEA's Measures of Academic Progress (MAP) tests are designed to be vertically-scaled assessments of math and reading achievement, which facilitates an examination of student growth across grades. As a result, some of the challenges documented by Quinn (2014) in deciding how to model the "learning gain" outcome are of less concern in this dataset. In addition, the NWEA data comes from the 2008 through 2016's post-accountability era—a time in which it is at least conceivable that the dynamics of access to quality schooling have changed. We also implement a set of multi-level models that we think connect more clearly to the central research questions in this domain: The coefficients ("fixed effects" in the language of HLM) correspond to school-year gains and summer losses, while the variance components allow us to characterize a plausible range of gain/losses one should expect across students during those periods. These variance components connect directly to our primary research question: The larger the variation in summer losses across students, relative to the school year gains, the more summers are the time when end-of-school achievement disparities arise.

Data and Sample

NWEA Data

The data for this study comes from the NWEA's MAP assessment. The dataset contains math and reading scores based on a computer adaptive test designed to serve as part of a formative, benchmarking data system, purchased by about 7500 districts across all 50 states in the U.S. The MAP assessment is used as a supplementary tool to aid schools' in improving their instruction and meeting students' needs, not as the high-stakes test of record. Because the MAP assessment is intended to monitor students' progress throughout the school year, it is administered in both the fall and the spring.[2] The MAP test is scored using a vertical and interval scale, which the NWEA calls the RIT scale. In theory, the vertical scale allows comparisons of student learning across grades and over time, while the interval scale ensures that a unit increase in a student's score represents the same learning gain across the entire distribution. The vertically-scaled nature of this outcome data is essential to our ability to examine differences in achievement disparities as students move through grade levels. However, it is worth noting that vertical scaling is difficult to

[2] It is also administered in the winter by some districts, however that data is not used in the current analysis.

achieve and hard to verify (Briggs, 2013; Briggs & Weeks, 2009). Therefore, our findings regarding changes across grades rely on assuming that NWEA's vertical scale is valid. However, much of the paper concerns itself with comparing learning gains in the same grade (a given school-year relative to the subsequent summer).

The full dataset used for the current study comes from 7,685 U.S school districts that administered the MAP assessment during the nine years between 2008 and 2016. Different districts opt to administer the MAP in different grades, however the NWEA full data includes 203,234,153 test scores for 17,955,222 million students who took a test between grades K and 11th grade. The dataset includes students' race and gender, their math and reading MAP scores, number of items attempted and correctly answered, duration of the test, grade of enrollment, and the date of test administration. The file does not include indicators for whether the student is an English Language Learner, belongs to the federal Free- and Reduced-Price Lunch program, or receives special education services.

Adjustments to NWEA RIT Scores

Students do not take MAP tests *exactly* on the first and last day of school—in fact, students often take these tests 3 to 6 weeks before/after the school year starts/ends. As a result, some of the time between the spring and fall administrations of the test—what one would mislabel as summer time—is actually spent in school. While the NWEA dataset does include the test date, crucially, it does not include school-year start or end dates to know exactly how much this occurs.

We therefore conducted a large-scale data collection effort to record the start- and end-date in every district in a subset of 11 states with the greatest use of MAP assessments. We found 23,223 school year start dates and 20,807 school year end dates—about 77 percent of the district-year calendar dates in those 11 states from 2008 to 2015. In later years, NWEA also began to collect school-year start and end dates. Together, these efforts allowed us to collect actual calendar start/end dates for 50.3 percent of the observed school-years for the entire NWEA dataset. Based on that data, we also extrapolate likely dates for other districts. We then use this calendar data to make a linear projection of each students' score on the first and last day of the school year. For more information about this process, including a description of our approach to collecting this data, the percent of actual dates recovered, our extrapolation process, our score projection process, and how study results differ when using observed scores instead of projected scores, see Appendix A. For fall ELA scores, the correlation between observed and projected RIT scores is 0.996, with an RMSE of 2.3 points.

Figure 2 illustrates how even small changes in estimated scores using projection methods could have a large impact on estimating summer learning rates.[3] Figure 2 presents two hypothetical students as they progress through school between January 2008 and January 2012. The first student's observed scores—and their test dates—are shown in orange. In dashed green, we project the student's achievement scores linearly based on their school-year learning rate. The green line connects the student's projected achievement on the last day of school to the projected achievement on the first day of school after that summer. In some cases, the summer learning rate estimated in the absence of school calendar information is positive, while learning rate is actually negative once the projections are used. The results are similar for the second student (red solid= observed scores, blue dashed= projected scores). The main takeaway here is that the linear projection process—though it produces scores strongly correlated with the observed scores—could have a profound

[3] Because the summer learning rate is estimated off of just two points—the first and last day of school—the slope between those points is quite sensitive to even minor adjustments. Note that the method we describe assumes that students learn just as much on days in May as the do in, say, February. While there is some evidence that learning rates are relatively linearly within the school year (Fitzpatrick, Grissmer, & Hastedt, 2011), there are also reasons to question this assumption, especially given anecdotal reports that the intensity of school activities slows after spring standardized test are given.

impact on the estimated summer learning gain/loss. Throughout this paper, we therefore use the projected RIT scores in favor of the observed RIT scores. However, in Appendix A, we reconduct the analyses using observed scores in place of projected scores and replicate the figures in this paper that capture the main findings.

Analytic Sample

For the current analysis, we first restrict the NWEA sample to students observed in grades 1 through 8 (because these are the grades with most complete coverage) and to the 89 percent of those students who neither repeat or skip grades. In our preferred models, we also restrict the sample to a "balanced panel"—that is, the subset of students who possess test scores for the full grade range being included in the model. For instance, if we examine test score patterns between 1st through 5th grade in a given model, only students who have both fall and spring test scores in every grade between 1st and 5th grade (that is, a full vector of all 10 reading test scores) will be included in the sample. While this is quite restrictive sample limitation, it ensures that our findings cannot be conflated with compositional changes from one time point to the next. In Appendix B, we replicate our primary findings on a less restrictive sample by running models with only 3 consecutive grades at a time (e.g., grades K through 2, grades 3 through 5, etc.). In these models, more students are included because the vector of required test scores is much shorter. These two samples have different advantages in terms of internal and external validity. Ultimately, however, results are relatively consistent (see Appendix B).

In Table 1, we compare the demographic descriptives for the students, schools, and districts from 4 groups: The population of U.S. public schools (from Common Core of Data), the entire population of NWEA test takers, and the subset of students who meet the *less* restrictive inclusion criteria (for Appendix B), and the *more* restrictive inclusion criteria for our preferred results. See

Table 1 (for simplicity, we conduct this comparison in the 2011-2012 school year). First, recall that that a student-level indicator of free/reduced-price lunch (FRPL) status is not available in the NWEA dataset. However, at the school level, the mean percent of students in a school who are FRPL-eligible is very similar across the four groups: 50 percent both nationally and in NWEA universe of schools, 48 percent in the larger Appendix B sample, and 51 percent in the more restrictive, primary analytic sample. In many ways, the NWEA sample reflects the U.S. public school population. For instance, it is similar in terms of percentage of students identified as Black, Asian, White, and male. In addition, the majority of U.S. public schools are in rural geographic codes, followed by suburban and rural geographies, and this ordering also holds in NWEA. Many of the district characteristics are also quite similar.

To consider potential (limitations of) generalizability, we point out that the largest differences between the U.S. public school population and the NWEA universe are that (a) the NWEA sample has a lower percentage of Hispanic students, (b) the average NWEA school has somewhat smaller mean enrollment, and (c) the NWEA districts tend to have more schools in them, have a lower percentage of FRPL students, and are less likely to be rural. What is also of note, however, is the sheer number of students in the NWEA universe in 2012 alone. NWEA students make up more than 11 percent of the entire K-12 public school population in 2012. NWEA data is available in nearly 37 percent of all U.S. public schools and in over half of all districts. This population is large enough to be of interest in its own right.

Finally, we examine how the analytic sample limitations affect the characteristics of the NWEA students included in the models (compare the right three columns of Table 1[4]). The final column reflects the requirements for inclusion in the balanced panel. For the most part, the analytic

[4] The analytic samples in this paper are first limited to NWEA students observed in grades 1 through 8, hence the large drop in sample size between the full NWEA sample and the Appendix B analytic sample.

restrictions do not dramatically alter the descriptive profile of included NWEA students, schools or districts. However, the primary analytic sample has a higher percentage of white students than the NWEA full dataset (60 percent versus 53 percent), and the schools tend to be a little smaller (mean enrollment of 391 versus 486) and are less likely to be suburban.

Methods

We adopt a multilevel model to estimate an individual learning trajectory for each student as they progress through sequential school years and summers. We then look across students to estimate how much students tend to gain, on *average*, during the school year versus what they typically lose during the summer. A multilevel modeling approach also allows us to estimate the *variation* in these gains/losses across students. Because estimates from such a model are empirically Bayes shrunk, we believe these are conservative estimates of student-level variances and therefore preferable to simply calculating the raw standard deviation of summer gains, which almost certainly reflect a great deal of measurement error (Raudenbush & Bryk, 2002).

Longitudinal Multi-Level Models

We use a two- level random effects (hierarchical) model, in which the outcome of interest is a test score, $Score_{ti}$, for student *i* at grade-semester *t*. In our preferred models, we separately model scores in 1st through 5th grade (students included here must have all 10 math score outcomes) and then in 5th through 8th grade[5] (again, students must have all 6 test scores in these grades). For brevity, we present the model (Eq. 1) for math scores from grade 6 through grade 8. These six repeated observations (L1) are nested within students (L2):

[5] We include 5th grade in both panels to informally check how similar 5th grade estimates are across the models.

Level One: Repeated Observations of Students (i) across Grade-Sems (t)

$$Score_{ti} = \pi_{0i} + \pi_{1i}(schyr6_{ti}) + \pi_{2i}(sumaf6_{ti}) + \pi_{3i}(schyr7_{ti}) + \pi_{4i}(sumaf7_{ti}) + \pi_{5i}(schyr8_{ti}) + \pi_{6i}(sumaf8_{ti}) + \varepsilon_{ti} \qquad \text{where } \varepsilon_{ti} \sim N_{iid}(0, \sigma)$$

Level Two: Students (i)

$$\pi_{0i} = \beta_{00} + r_{0i} \qquad \text{where } r_{0i} \sim N_{iid}(0, \tau_{0,0})$$
$$\pi_{1i} = \beta_{10} + r_{1i} \qquad \text{where } r_{1i} \sim N_{iid}(0, \tau_{1,1})$$
$$\vdots$$
$$\pi_{6i} = \beta_{60} + r_{6i} \qquad \text{where } r_{6i} \sim N_{iid}(0, \tau_{6,6}) \qquad \text{Eq (1)}$$

At L1, students' growth trajectories are modeled with a set of dummy variables—$schyr6_{ti}$, $sumaf6_{ti}$, $schyr7_{ti}$, $sumaf7_{ti}$, etc.—for each grade-semester. They are each coded as 1 if the observation occurred *on or after the ending timepoint for the period*.[6] This coding scheme may at first seem confusing, but it has the major advantage of giving the level-one coefficients intuitive meaning that now match the variable names: They represent an individual student *i*'s grade-specific school-year gain or grade-specific summer gain/loss. For example, π_{1i}—the coefficient on $schyr6_{ti}$—captures student i's 6th grade school-year learning gain. The coefficient on $sumaf6_{ti}$ captures student *i*'s summer after 6th grade gain/loss. These coefficients are now the very learning gains/losses we are interested in estimating for each student. We allow all of the level-one coefficients π_{0i} through π_{6i} to vary randomly at the student level, and we assume that the level-two errors (r_{0i} through r_{6i}) are normally distributed with a mean of zero and a constant variance given by $\tau_{0,0}$ through $\tau_{6,6}$. These models estimate the parameters we need to answer each of our research questions, in turn.[7]

---

[6] For example, $schyr6_{ti}$ takes a value of 1 at the end of 6th grade (i.e., grade six spring test score) and remains a 1 for all observations thereafter. And $sumaf6_{ti}$ takes a value of 1 at the end of the summer after 6th grade (i.e., grade seven fall test score) and remains a 1 for all observations after.

[7] The parameters are presented with a focus on their substantive meaning in the Results section, but for those interested in a more formal roadmap between research questions and parameters: For RQ1 concerning mean gains/losses, we focus on the β coefficients. For RQ2 concerning student-level variation in gains/losses, we interpret the $\boldsymbol{\tau}$ variance parameters. For RQ3 concerning whether the same students tend

Results

(RQ1) Average Students' School-Year vs. Summer Learning Gains/Losses across Grades

Throughout the Results section, we present findings both formally (i.e., point estimates in tables) and visually to make takeaways as tangible as possible. For instance, to address this first question, we present the β coefficients (or "fixed effects" in the language of HLM) in Table 2 (ELA) and Table 3 (math) because, substantively, they capture mean gains/losses in each grade and the summer after. These β coefficients are also graphed in Figure 3 as mean growth trajectories. To best contextualize the findings about summer experiences, we first set the stage with a discussion of the magnitude of school-year learning gains.

During school years. Beginning with ELA school-year gains (left column of Table 2), we find that students' school year learning gains are largest in the early grades and generally diminish over time. This is depicted in Figure 3 with blue, dashed lines. For instance, students gain on average 23.7 ELA test score points in 1st grade, 18.5 points in 2nd grade, 13.3 points in 3rd grade, and so on. By 8th grade, the average ELA learning gain on NWEA's RIT scale is just 4.4 points. We observe a very similar pattern for math (left column of Table 3). In all grade levels, the average student gains—as opposed to loses—ground during school years. This suggests that students accumulate knowledge over time during school years as measured by the NWEA MAP test.

During summers. The pattern of mean summer learning gains/losses—the β coefficients in the right column of Table 2 and Table 3—are shown as solid red lines in Figure 3. Summer estimates differ from school-year gains in two important ways. First, in both ELA and math, the summer coefficients between 1st and 8th grade are negative and tend to be smaller in magnitude.

---

to lose ground summer after summer (take, for example, the relationship between losses in the summers after 6th vs. 7th grade), we correlate students' empirical Bayes shrunken estimates of π_{2i} and π_{4i} (for this example, the covariance is $\boldsymbol{\tau_{2,4}}$[7]). For RQ 4, we make use of the student-level Bayes shrunken residuals.

For instance, the average ELA loss in the summer after 1st grade is -6.6 test score points, -3.9 in summer after 2nd, -3.4 in the summer after 3rd, and falls to a low of -0.9 by the start of grade 8. In math, the mean summer learning estimates are also negative and of similar magnitude. An implication here is that, depending on grade, the average student loses between 17 and 28 percent of their school year ELA gains (a 9-month period) during the following summer (a 3-month period). In math, the relative losses are a little larger: The average student loses between 25 and 34 percent of each school year gain during the following summer.

The second way in which summer estimates differ from their school-year counterparts is that the magnitude of mean summer learning losses does not decrease over time to the same degree as school year learning. Put differently, although mean school year gains in ELA fall from 23.7 to 4.4 across grades, mean summer losses stay within a tighter range of -6.6 to -0.9.

Turning to the visual representation of these findings in Figure 3, we consistently see a zig-zag pattern at every grade level, though the intensity of gains/losses flattens at higher grades. These results generally confirm the notion that summers can be characterized as a time when, on average, student learning slows or slides backwards. These findings differ somewhat from previous work in the apparent magnitude of the phenomenon. For instance, Alexander, Entwisle, and Olson (2001) found a strong seasonal pattern to learning gains in their Baltimore sample from the 1980s. However, they find that average learning rates during elementary grade summers slow dramatically but do not actually flatten or become negative. In other words, they find that learning during the summer slows but remains slightly positive. Downey et al. (2004), on the other hand, use more recent ECKS-K:1999 data and find that students neither gain nor lose ground during the summer after K (i.e., flat). While all three data sources—BSS, (1980s), ECLS-K (1999) and the current data (2008-2016)—show evidence that learning rates differ between summers and school years, ours is the only one which shows consistent evidence of mean learning *losses* during the summer at every grade level.

(RQ2) Variation in Students' School-Year vs. Summer Learning Gains/Losses, by Grade

It is important to recognize that the trends illustrated in Figure 3 only tell us one part of the story: the seasonal learning patterns for the *average* student. However, achievement disparities are driven by *differential* learning patterns, and so we now focus on how students vary on both school-year and summer learning gains/losses. We are particularly interested in determining whether student growth trajectories vary more during school years or summers.

During school years. We begin with an examination of variability in school-year learning across students. Also reported in the first column of Table 2 (ELA) and Table 3 (math) are the estimated standard deviations (SDs) of learning gains/losses across students in and after each grade (i.e., the square root of the diagonal elements of the tau matrix). As an example, while we saw before that the average student gain in 1st grade ELA was 23.7 points, students also typically differ from this mean by 9.7 points, suggesting a notable range across students in 1st grade school years gains. To provide context for the magnitude of this variability, under the assumption of normality across students (Raudenbush & Bryk, 2002), we construct a 95 percent plausible value range ("PVR") for learning gains across students. These are also reported in Table 2 (ELA) and Table 3 (math) beneath the corresponding student SD. To follow up with the example of ELA gains in 1st grade, we expect that 95 percent of students would have an average learning gain between 4.4 and 42.7 ELA test score points. Therefore, in 1st grade, students at the high end of the PVR gain about 80 percent more than the average student.

Estimates of the standard deviation of school-year learning gains across students are relatively consistent across school years and subjects, generally in the range of 6 to 10 test score points. In grades that exhibit smaller *average* school year gains, this variation implies larger discrepancies across students. For instance, in 8th grade when average growth is only 4.4 test score points during the school year, we see a 95 percent PVR across students of -7.0 to +15.9 points. Here, students at the top of this PVR will experience nearly four times larger gains than the average student. Students at the low end of that same PVR, however, are actually *losing* ground during 8th grade.

To juxtapose mean gains/losses with variation around them, we calculate the ratio of the variation (SD) across students for each learning gain to the mean learning gain. Larger ratios indicate greater variability, relative to the mean gain. In 1st grade ELA, that ratio is about 0.41 (9.7 over 23.7), indicating that the SD is a little less than half the size of the mean gain. In ELA, that ratio grows slowly across grades and reaches 1.3 in grade 8 (that is, the SD is now about 30 percent larger than the mean). The ratio also increases across grades in math, but less dramatically from 0.40 in 1st grade to 0.91 in 8th grade. However, the fact that the relative variability in learning gains grows as students progress through school may suggest that inequities in achievement accumulate to some extent during school years as students who are underprepared are being left further and further behind with each successive grade.

During summers. While the variability in school-year patterns are interesting in and of themselves, our main interest lies in whether the summer gains/losses vary more than gains in the school year periods. This has direct implications for our understanding of when discrepancies in student achievement arise across the course of students' school-age years. Turning to the second columns of Table 2 (ELA) and Table 3 (math), we see that the SD for a given summer tends to be a little smaller than the SD in the preceding school year (with the exception of 1st grade). For instance, in 3rd grade math, the SD is 6.6 in the school year and 3.6 in the following summer.

However, in a relative sense, the summer SDs are *much* larger with respect to means. In ELA, the SD-to-mean ratios described above are much larger in summers, ranging from 1.4 to as high as 5.2. A ratio of 5.2 indicates that the SD is over five times larger than the mean (recall that the largest such ratio during a school year was 1.4). In math, we also see that summer ratios, which range from 0.8 to 2.3 are larger than school-year ratios (ranging from 0.40 to 0.91). Keep in mind that this larger summer variation is arising in a comparatively shorter time (9 versus 6 months).

The PVRs are large for summer learning loss. Take 2nd grade math as an example: Summer learning loss in grade 2 for math (second column of Table 3) ranges from -16.3 to +6.8. While students at the top of that PVR are gaining, during the summer, another 32 percent of average growth from the preceding 2nd grade school year (6.8 over 18.6), students at the bottom of the PVR will *lose during the summer just as much as the typical student gained in 2nd grade*. Looking across all grades in ELA, we find that students at the top of the summer loss PVR will gain during the summer between 45 to 154 percent of the mean growth in the preceding grade (12 to 86 percent for math). However, students at the bottom of the summer loss PVR will *lose* during the summer between 93 to 194 percent of the mean growth in the preceding grade (73 to 136 percent for math). In sum, some students experience accelerated learning during the summer, while others lose nearly all of their gains from the preceding school year.

The takeaways for RQ2 are also illustrated visually in Figure 4 (ELA) and Figure 5 (math), wherein we present box plots of individual students' empirical Bayes estimated learning gains and losses in each school year and summer. These concisely capture the essence of what is presented in the tables: Larger gains during school years that diminish across grades, smaller average losses during summers that are more consistent is magnitude, but real variability around typical gains/losses. In Appendix B, we replicate Figure 4 (ELA) and Figure 5 (math) using results from models using a shorter three-grade increment. Though the data is sparser before 1st and after 9th grade, we include those grades in Appendix B.

In sum, students certainly appear to vary in terms of how much they learn during the school year, but most students tend to exhibit some growth in test scores while in school. However, the picture in the summer is quite different. While our results re-document the known *mean* summer learning loss phenomenon, this finding obscures a more problematic pattern: For some as-yet unknown reason, certain students can gain at a faster rate in the summer than the mean rate in the preceding school year, while other students could lose most of what is typically gained.

(RQ3) Student-Level Correlation of Summer Gains/Losses across Summers

To this point, we have highlighted important variability in summer learning patterns across students. However, if that phenomenon occurs to students random—that is, a student might gain in one summer and then randomly lose in the next—then the contribution of summer learning loss to end-of-school achievement *disparities* would be limited. However, if the same students tend to experience losses summer after summer, while others gain summer after summer, it would lead to a more dramatic "fanning out" of student outcomes as they progress through school. We are particularly concerned if the students who exhibit the greatest summer losses also tend to be from historically marginalized student populations (a question outside the scope of the current paper and difficult to tackle with the data at hand).

To explore this question empirically, we examine from our multilevel models the estimated covariances of students' summer losses across grades.[8] Table 4 (ELA) and Table 5 (math) present

---

[8] Returning briefly to Equation (1) for a concrete example, consider the covariance of the π_2i's (student i's estimated change in the summer after 6th grade) with the π_4i's (in summer after 7th grade). That covariance ($\tau\_4,6$) captures the extent to which students who lose ground in one summer tend to be the same ones who lose ground in the next summer. Like the variances presented earlier, these estimated covariances have been empirically Bayes-shrunk to account to some degree of measurement error.

these covariances (expressed in correlations). Positive correlations suggest that summer gains/losses accrue to the same students across grades in a way that would contribute to the widening of end-of-schooling student outcomes. Correlations near zero would suggest gains/losses occur somewhat randomly. In ELA, all correlations are positive (between 0.12 and 0.65), and most are substantively large. All of these correlations are also positive in math, ranging between 0.10 and 0.65. This suggests that students' experiences across summers tend to be similar. Those who lose ground one summer are more likely to also lose in the next.

(RQ4) Proportion of Variation in Outcomes that Arises in Summers

Taken together, these three factors—(1) slightly negative mean summer losses, (2) large variances in summer loss/gains, and (3) systematic gain/loss patterns across summers—imply that end-of-school achievement disparities arise partly during the summer. How large of a role do summers play? To consider this question, we begin by presenting a thought experiment designed to characterize the role of summers between grade 1 and 8. We imagine a hypothetical scenario in which all students enter 1st grade at the exact same achievement level, and all students experience the exact same (let's say, the mean) learning gain in each grade while school is in session. If there were no summer periods, all students in this scenario would end 8th grade with the same test score, because no variation in gains arises while in school. We now return to the results from our multilevel model to characterize three plausible student experiences during the summers following each grade: The typical gain among students in the top, middle, and bottom thirds of a given summer's gain/loss distribution.[9] We now illustrate these three levels of summer experiences in

[9] We split the distribution of student-specific, empirical Bayes shrunken summer learning gain/loss estimates into a top, middle, and bottom tercile and then calculate the mean learning gain within each of those terciles. We do this separately for residuals for each summer following a school year between first and 8th grade.

Figure 6 (ELA in top panel, math in bottom panel), while assuming school year gains are always equal (i.e., parallel slopes of dashed blue lines fall to spring).

This figure shows how the differences in summer experiences *by themselves* could lead to sizeable achievement over time. In ELA, the spread in test scores at the end of 8th grade is from about 185 to 255 test score points (and about 200 to 265 in math)—around 2.5 standard deviations of spring 8th grade RIT scores. This thought experiment illustrates the idea that, even in an ideal world where school inequities could be eliminated, achievement gaps would arise simply because of the summer break. The "fanning out" of achievement during these school-age years would need to be addressed in large part with respect to summer experiences.

Finally, to put a point on the question of the summer's role in achievement disparities, we calculate for each student the sum of all fluctuations in their test scores during a panel (here, from the start of 1st to the end of 5th grade) and then calculate what percentage of those fluctuations arose during summers. For a student who experiences no change in their scores from the start to the end of the summers (i.e., always flat slopes in the summers), this proportion would be close to zero. For a student who year after year exhibits summer learning loss (or gains), this proportion would be larger. In Figure 7, we present the distribution of those proportions across students. On average, summer gains/losses account for 19.4 percent of students' test score changes between 1st and 5th grade occur during the summer (19.3 for math). However, for some students, summer fluctuations account for much more—even upwards of 30 percent—of where they end up.

Conclusion

Reflections on Findings

In this paper, we conduct a thorough exploration of the seasonality of learning from a dataset covering nearly 18 million students in 2008 through 2016 across all 50 states. We focus on characterizing the degree of variability in students' summer experiences and the role of summers in contributing to end-of-school achievement disparities. We find that students do indeed tend to lose ground during the summer period in both math and ELA. We add to the existing research by also estimating the variance across students in summer learning loss: For instance, in the summer after 2nd grade, the 95 percent plausible value range indicates that some students will *lose* as much as 16.3 test score points in math during summer, while other students could *gain* up to 6.8 test score points (relative to a mean loss of 4.8 points). Students also exhibit significant variance in school year learning, however the lower bounds of the 95 percent plausible value ranges during the school year tend to be much closer to zero. This means that, while some students learn more than others during the school year, most students are moving in the same direction—that is, making learning *gains*—while school is in session.

The same cannot be said for summers. During the summer, a little more than half of students exhibit summer learning losses, while the others exhibit summer learning gains. It is clear that the summer period is a particularly variable time for students. We find that many students can in fact maintain average school-year learning rates during the summer in the absence of formal schooling. Other students, however, will lose nearly as much as what is typically gained in the preceding school year. This remarkable variability in summer learning rates appears to be a strong contributor to widening achievement disparities as students move through school. On average, about 19 percent of the fluctuations in students' achievement between 1st and 5th grade occurs during the summer. However, for some students, summers account for as much as 30 percent. This is a particularly outsized influence given that the summer makes up only about a quarter of each calendar year.

Study Limitations

First and foremost, the NWEA dataset does not include some key student variables that one would want available to more deeply understand the between-student variance in summer learning rates (e.g., FRPL status, language status, special education status). Moreover, a key component of the learning equation is unavailable to us in the NWEA data—links to individual teachers. In addition, the current study rests on the assumption that NWEA's RIT scores are a suitably valid measure of student math and reading skills in both the fall and spring and over time (i.e., the vertical scaling). NWEA's MAP test is a formative assessment without stakes, and it is not entirely clear that there are incentives in place for students and teachers to take it equally seriously in the fall and spring. Students tend to spend slightly less time on the fall tests than their spring tests. One would be concerned if this signals that students do not put forward as much effort on their fall assessments, thus making summer learning losses appear larger than they actually are. We believe that the difference in time spent is not large (about 10 additional seconds per item, on average), and we find that controlling for time spent on test affects the results very little. In addition, most of the analyses herein do not rely on making direct comparisons across distal grades, thus reducing reliance on vertical scaling properties for these particular inferences. That said, the findings herein should be considered with this caveat in mind.

Implications

Our results do show that achievement disparities widen during school years. As such, we should of course continue to develop policies that change how students experience schools, particularly around issues of access. On the other hand, we find that—even in an ideal scenario in which students all learn the same amount during the school year—the time spent out of school in summer break, by itself, gives rise to much of the dramatic spread of achievement outcomes, on the order of several standard deviations.

A natural policy idea, then, is to extend the school year to reduce summer atrophy and minimize opportunities for this divergence to occur. However existing research on year-round school calendars does not indicate that SLL is mitigated by these schedules (Graves, 2011; McMullen & Rouse, 2012). It is possible that year-round calendars implemented to address over-crowding (a common impetus) may have different impacts on learning than year-round calendars implemented explicitly to reduce SLL, but to our knowledge this hypothesis has not been tested.

Another policy lever might be focus on programs that bridge the gap between May and August like summer school. The causal evaluation of summer school is often fraught, given the non-random selection of who is required to enroll and known issues around low attendance (especially in higher grades). Yet there is growing evidence that summer interventions can help mitigate students' SLL (Kim & Quinn, 2013; McCombs et al., 2012; McCombs et al., 2015). For instance, seven New Mexico school districts randomized early grade children in low-income schools into an ambitious (and presumably expensive) summer program called K-3+, that essentially amounted to a full-blown extension of the typical school year for much of the summer period. Early results from the experimental study indicated that children assigned to K-3+ exhibited stronger literacy outcomes across four domains of the Woodcock Johnson achievement assessment (Cann, Karakaplan, Lubke, & Rowland, 2015).

Finally, it remains an open—but important—question whether schools can or should be held accountable for students' summer learning experiences (for a discussion, see McEachin & Atteberry, 2017). On the one hand, schools do make some efforts to provide curricular guidance for students after the school year ends (e.g., summer reading books). On the other hand, if some schools serve students who are systematically more likely to lose ground during the summer, then those schools face a quite different challenge at the start of each school to bring their students back to where they left off when school ended the previous spring. Traditional statewide testing systems would not be able to detect such a pattern, and school accountability measures based on spring-only scores essentially conflate schools' impacts on both school year and summer learning.

Next Steps for SLL

The objective of this paper is to document the magnitude of a social problem—the role of summers in the "fanning out" of student achievement outcomes as students move through school. While we can conclude that this happens and to what extent, the current dataset is not well-positioned to understand *why* summer learning patterns are so varied across students. Though it is an important first step to know when inequality arises and how unequal the learning patterns are, the obvious next question is: What accounts for that variation?

In some sense we have reached a precipice on SLL research. It seems clear that summers play a key role in outcome inequality and that the range of students' summer learning patterns is sizeable. This variability may fall partly along racial and socioeconomic lines (Burkam et al., 2004; Entwisle & Alexander, 1992; Gershenson, 2013; Quinn, 2014). However, demographic factors do not account for much of the story here. In an insightful SLL study by Burkam et al. (2004) using ECLS-K:1999 data, the authors leverage the parent surveys of children's home and summer activities, in conjunction with student gender, racial, and socio-economic demographics—that is, most of the first-order candidates for explaining variability. However, they can explain only about 13 percent of the variance in learning gains in the summer after K. New research is needed to reconcile the fact that summer learning differs dramatically from child to child, but to date we have little insight into what drives that variation.

Tables

Table 1. Descriptive Statistics in the Nation, in Full Dataset, in Analytic Sample in 2011-12

Level	Statistic	All U.S. Public Schools	Full NWEA Dataset	App B: Analytic Sample	Primary: Analytic Sample
Student-Level	% FRPL	0.46	n/a	n/a	n/a
	% Black	0.16	0.12	0.11	0.12
	% Hispanic	0.24	0.12	0.12	0.09
	% Asian	0.05	0.04	0.04	0.03
	% White	0.52	0.53	0.58	0.60
	% Male	0.51	0.51	0.50	0.50
	Total N of Students in 2012	*49,256,120*	*5,469,366*	*1,892,098*	*260,037*
School-Level	Average Enrollment	532	486	432	391
	Mean % FRPL	0.50	0.50	0.48	0.51
	Mean % Black	0.15	0.15	0.12	0.17
	Mean % Hispanic	0.21	0.17	0.16	0.13
	Mean % Asian	0.04	0.03	0.03	0.03
	Mean % White	0.56	0.60	0.63	0.60
	% of Schools in Urban Locale	0.25	0.23	0.22	0.26
	% of Schools in Suburban Locale	0.32	0.24	0.25	0.16
	% of Schools in Rural Locale	0.43	0.32	0.37	0.47
	Total N of Schools in 2012	*89,648*	*32,755*	*10,533*	*1,440*
District-Level	Average N of Schools in District	7	9.1	8.8	12.9
	Mean % FRPL	0.45	0.36	0.35	0.34
	Mean % Black	0.07	0.07	0.06	0.05
	Mean % Hispanic	0.13	0.12	0.11	0.11
	Mean % Asian	0.02	0.02	0.02	0.02
	Mean % White	0.73	0.76	0.78	0.78
	Mean % Male	0.52	0.52	0.51	0.51
	Mean Stu:Tch Ratio	14.5	14.8	14.4	13.9
	% of Districts in Urban Locale	0.06	0.04	0.03	0.05
	% of Districts in Suburban Locale	0.29	0.20	0.19	0.17
	% of Districts in Rural Locale	0.63	0.44	0.51	0.51
	Total N of Districts in 2012	*13,273*	*7,437*	*3,242*	*1,093*

FN: Data for U.S. public school population comes from the NCES Common Core of Data and has been restricted to public schools (https://nces.ed.gov/ccd/). FRPL status is not available at the student level in the NWEA data. The Appendix B sample includes more NWEA students because it does not require students to have as long of a panel of available test scores to be included. The primary analytic sample used in the main narrative requires students to have up to ten available test scores in a row without missing data.

Table 2. ELA: School-Year & Summer Learning Gains/Losses, Student-Level Standard Deviations, 95% Plausible Value Ranges across Students

		Model-Based Estimates		Post-Hoc Statistics for Given Grade					
		Gains/ Losses during the School Year	Gains/ Losses during the Following Summer	Means: % of Schyr Gain Lost in Summer	% More Gained @ Top of PVR in Schyr	Schyr: Ratio of SD to Mean Gain	Summer: Ratio of SD to Mean Gain	Summer: % of SY Gained @ Top of PVR	Summer: % of SY Lost @ Low of PVR
Grade 1	*coeff (beta)*	23.7 ***	-6.6 ***	28%	80%	0.41	1.6	55%	114%
	(se of beta)	(0.05)	(0.05)						
	stud sd (tau)	9.7 ***	10.4 ***						
	(stud 95% PVR)	(4.6 to 42.7)	(-26.9 to 13.7)						
Grade 2	*coeff (beta)*	18.5 ***	-3.9 ***	21%	109%	0.56	1.7	49%	93%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	10.3 ***	6.8 ***						
	(stud 95% PVR)	(-1.6 to 38.7)	(-17.2 to 9.3)						
Grade 3	*coeff (beta)*	13.3 ***	-3.4 ***	26%	119%	0.61	1.4	45%	98%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	8.1 ***	4.9 ***						
	(stud 95% PVR)	(-2.6 to 29.3)	(-13.0 to 6.3)						
Grade 4	*coeff (beta)*	10.1 ***	-2.6 ***	26%	132%	0.67	1.8	59%	118%
	(se of beta)	(0.04)	(0.04)						
	stud sd (tau)	6.8 ***	4.7 ***						
	(stud 95% PVR)	(-3.2 to 23.4)	(-11.9 to 6.7)						
Grade 5	*coeff (beta)*	7.8 ***	-2.2 ***	28%	204%	1.04	2.5	103%	169%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	8.1 ***	5.6 ***						
	(stud 95% PVR)	(-8.1 to 23.8)	(-13.2 to 8.8)						
Grade 6	*coeff (beta)*	6.4 ***	-1.6 ***	25%	236%	1.20	3.3	125%	186%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	7.7 ***	5.3 ***						
	(stud 95% PVR)	(-8.7 to 21.4)	(-11.9 to 8.8)						
Grade 7	*coeff (beta)*	5.2 ***	-0.9 ***	17%	275%	1.40	5.2	154%	194%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	7.3 ***	4.7 ***						
	(stud 95% PVR)	(-9.1 to 19.6)	(-10.1 to 8.4)						
Grade 8	*coeff (beta)*	4.4 ***		n/a	258%	1.32	n/a	n/a	n/a
	(se of beta)	(0.04)							
	stud sd (tau)	5.8 ***							
	(stud 95% PVR)	(-7.0 to 15.9)							

Table 3. Math: School-Year & Summer Learning Gains/Losses, Student-Level Standard Deviations, 95% Plausible Value Ranges across Students

		Model-Based Estimates		Post-Hoc Statistics for Given Grade					
		Gains/ Losses during the School Year	Gains/ Losses during the Following Summer	Means: % of Schyr Gain Lost in Summer	% More Gained @ Top of PVR in Schyr	Schyr: Ratio of SD to Mean Gain	Summer: Ratio of SD to Mean Gain	Summer: % of SY Gained @ Top of PVR	Summer: % of SY Lost @ Low of PVR
Grade 1	*coeff (beta)*	24.0 ***	-6.4 ***	27%	91%	0.46	1.7	58%	114%
	(se of beta)	(0.05)	(0.05)						
	stud sd (tau)	11.1 ***	10.7 ***						
	(stud 95% PVR)	(2.2 to 45.9)	(-27.4 to 14.6)						
Grade 2	*coeff (beta)*	18.6 ***	-4.8 ***	26%	92%	0.47	1.2	32%	88%
	(se of beta)	(0.04)	(0.04)						
	stud sd (tau)	8.7 ***	5.9 ***						
	(stud 95% PVR)	(1.6 to 35.6)	(-16.3 to 6.8)						
Grade 3	*coeff (beta)*	16.5 ***	-4.6 ***	28%	78%	0.40	0.8	12%	73%
	(se of beta)	(0.04)	(0.03)						
	stud sd (tau)	6.6 ***	3.7 ***						
	(stud 95% PVR)	(3.6 to 29.4)	(-12.0 to 2.7)						
Grade 4	*coeff (beta)*	14.2 ***	-4.3 ***	30%	86%	0.44	1.1	28%	96%
	(se of beta)	(0.04)	(0.03)						
	stud sd (tau)	6.2 ***	4.7 ***						
	(stud 95% PVR)	(2.0 to 26.3)	(-13.6 to 4.9)						
Grade 5	*coeff (beta)*	11.7 ***	-4.0 ***	34%	136%	0.69	1.3	51%	121%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	8.1 ***	5.2 ***						
	(stud 95% PVR)	(-4.2 to 27.5)	(-14.2 to 6.2)						
Grade 6	*coeff (beta)*	9.8 ***	-2.7 ***	28%	144%	0.73	1.8	61%	127%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	7.2 ***	4.9 ***						
	(stud 95% PVR)	(-4.4 to 23.9)	(-12.4 to 6.9)						
Grade 7	*coeff (beta)*	8.1 ***	-2.0 ***	25%	179%	0.91	2.3	86%	136%
	(se of beta)	(0.05)	(0.04)						
	stud sd (tau)	7.4 ***	4.6 ***						
	(stud 95% PVR)	(-6.4 to 22.6)	(-11.0 to 7.0)						
Grade 8	*coeff (beta)*	6.5 ***		n/a	163%	0.83	n/a	n/a	n/a
	(se of beta)	(0.04)							
	stud sd (tau)	5.4 ***							
	(stud 95% PVR)	(-4.1 to 17.2)							

Table 4. ELA: Student-Level Correlations of Estimated Summer Gains, across Grades

Summer After…	Grade 01	Grade 02	Grade 03	Grade 04	Grade 05	Grade 06	Grade 07
Grade 01	1.00						
Grade 02	0.65	1.00					
Grade 03	0.28	0.57	1.00				
Grade 04	0.20	0.25	0.56	1.00			
Grade 05					1.00		
Grade 06					0.54	1.00	
Grade 07					0.12	0.57	1.00

FN: The model is run separately on early grades and later grades. Because the panel is only 9 years long, very few (less than 1 percent) of students have all 19 test scores between first through eighth grades. We therefore cannot estimate correlations across these two models.

Table 5 Math: Correlation Matrix Across Students' Summers Losses

Summer After…	Grade 01	Grade 02	Grade 03	Grade 04	Grade 05	Grade 06	Grade 07
Grade 01	1.00						
Grade 02	0.65	1.00					
Grade 03	0.15	0.43	1.00				
Grade 04	0.09	0.15	0.49	1.00			
Grade 05					1.00		
Grade 06					0.42	1.00	
Grade 07					0.10	0.53	1.00

FN: The model is run separately on early grades and later grades. Because the panel is only 9 years long, very few (less than 1 percent) of students have all 19 test scores between first through eighth grades. We therefore cannot estimate correlations across these two models.

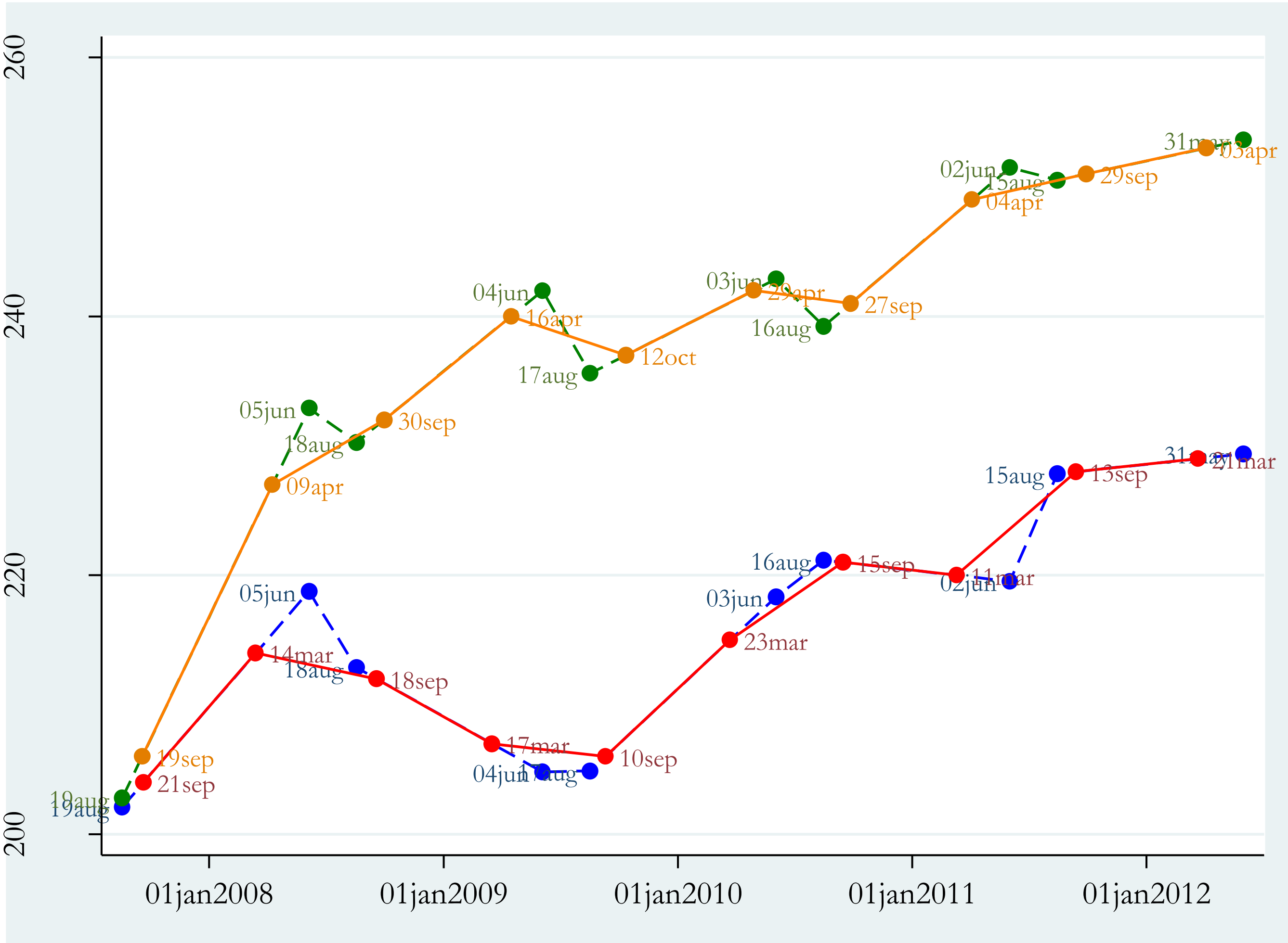
Figures

Figure 1. Compare Studies: Datasets, Data Years, Grades Included, Number of Students, Location

Authors	Publication Year	Dataset	Years Since Data Collected	Summers After Grades…	# of Students	Geography
Heyns	1978	unnamed	***47 years prior***	6th	3,000	Atlanta
Entwisle & Alexander	1992	BSS [1]	***2019 years prior***	1st, 2nd	790	Baltimore
Allinder, Fuchs, Fuchs, & Hamlett	1992	unnamed	At least 27 years prior	2nd, 3rd, 4th	275	2 rural schools in a midwest state
Alexander, Entwisle, & Olson	2001	BSS [1]	***32 years prior***	1st, 2nd, 3rd, 4th	678	Baltimore
Burkam, Ready, Lee, LoGerfo	2004	ECLS-K [2]	***20 years prior***	K	3664	nationally representative
Downey, von Hippel, & Broh	2004	ECLS-K [2]	***20 years prior***	K	~4000*	nationally representative
Borman, Benson, & Overman	2005	"Teach Baltimore"	***19 years prior***	K (2 cohorts)	303	Baltimore high-povery schools
Alexander, Entwisle, & Olson	2007	BSS [1]	***32 years prior***	1st, 2nd, 3rd, 4th	326	Baltimore
Downey, von Hippel, Hughes	2008	ECLS-K [2] + Census	***20 years prior***	K	4217	nationally representative
Benson & Borman	2010	ECLS-K [2] + Census	***20 years prior***	K	4180	nationally representative
Skibbe, Grimm, Bowles, & Morrison	2012	unnamed	***13 years prior***	Pre-k, K, 1st	383	1 suburban midwest town
Gershenson	2013	APSCC [3] / ATUS [4]	29 years prior / 9 years prior	n/a	628 / 23,348	California / US
Quinn	2014	ECLS-K [2]	***20 years prior***	K	3043	nationally representative
Current Study	--	***NWEA*** [5]	***3 years prior***	***1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th***	***18 million***	***Across all US states***

*FN: 1 BSS (Beginning School Study); 2 ECLS-K (Early Childhood Longitudinal Study: Kindergarten Class of 1999); 3 APSCC (Activity Pattern Survey of California Children; 4 ATUS (American Time Use Study. (time-diary surveys); 5 Northwest Evaluation Association-- current study. *This paper uses data from 17,212 students but does not count the subset of students for whom summer learning can be estimated. We know from other studies using ECLS-K this is about 4,000 students.*

Figure 2. Illustration of the Timeline for Observed and Projected RIT Test Scores



FN: Student 1: Observed scores in orange, projected scores in green. Student 2: Observed scores in red, projected scores in blue.

Figure 3. ELA and Math: Estimated Mean School Year Gains and Summer Losses

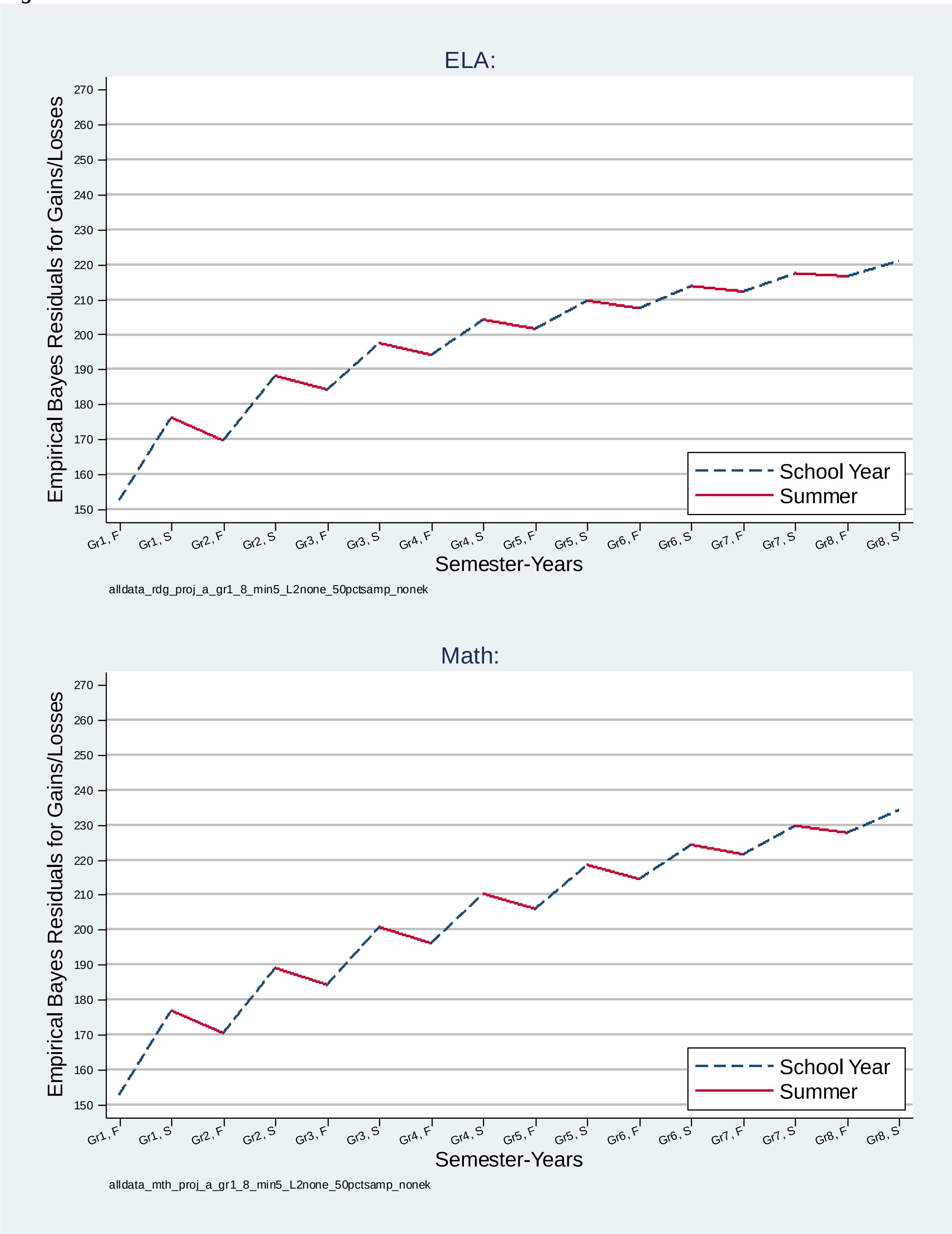


Figure 4. ELA: Boxplot of Students' Empirical Bayes Estimated Gains/Losses, across Grades

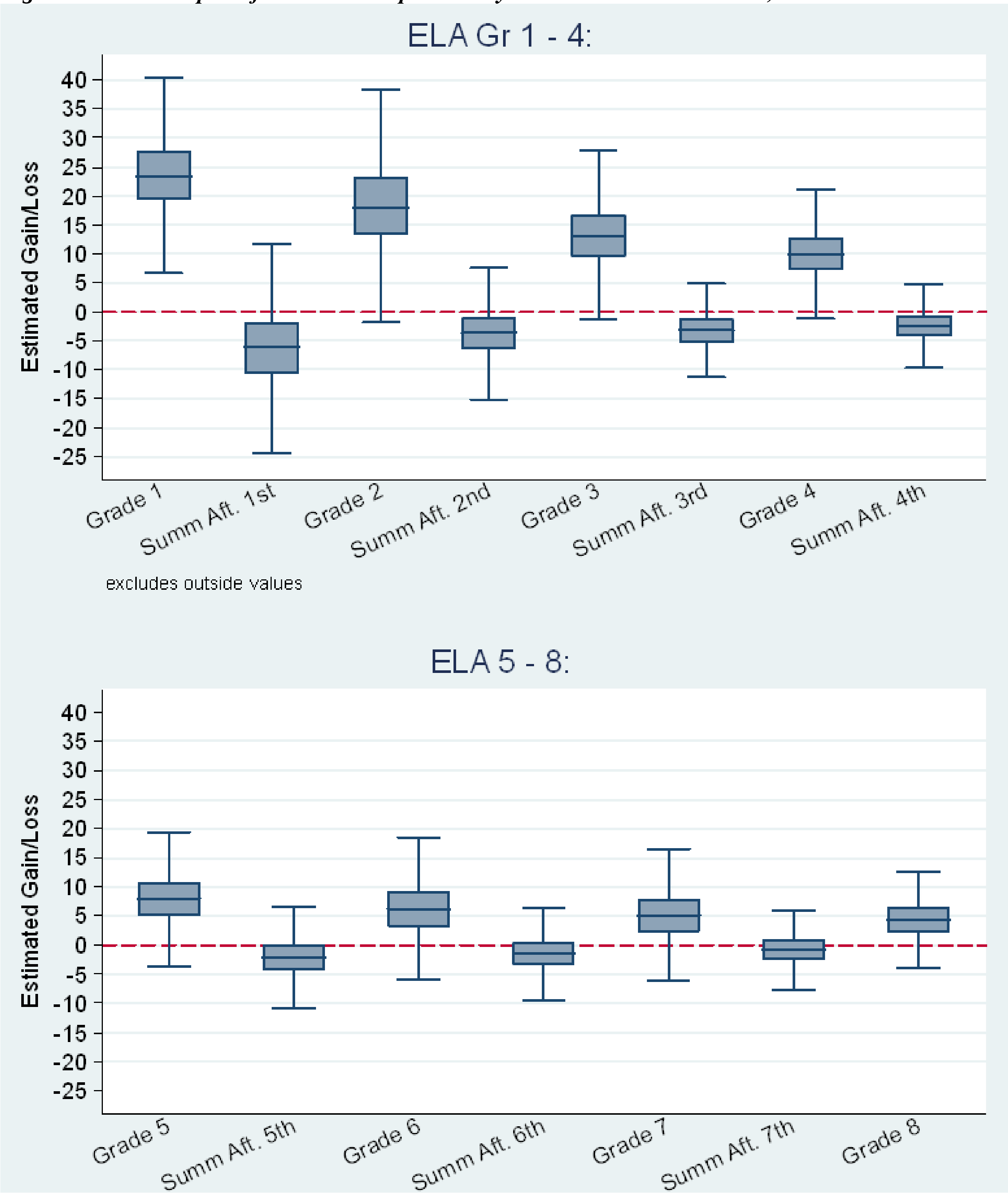


Figure 5. Math: Boxplot of Students' Empirical Bayes Estimated Gains/Losses, across Grades

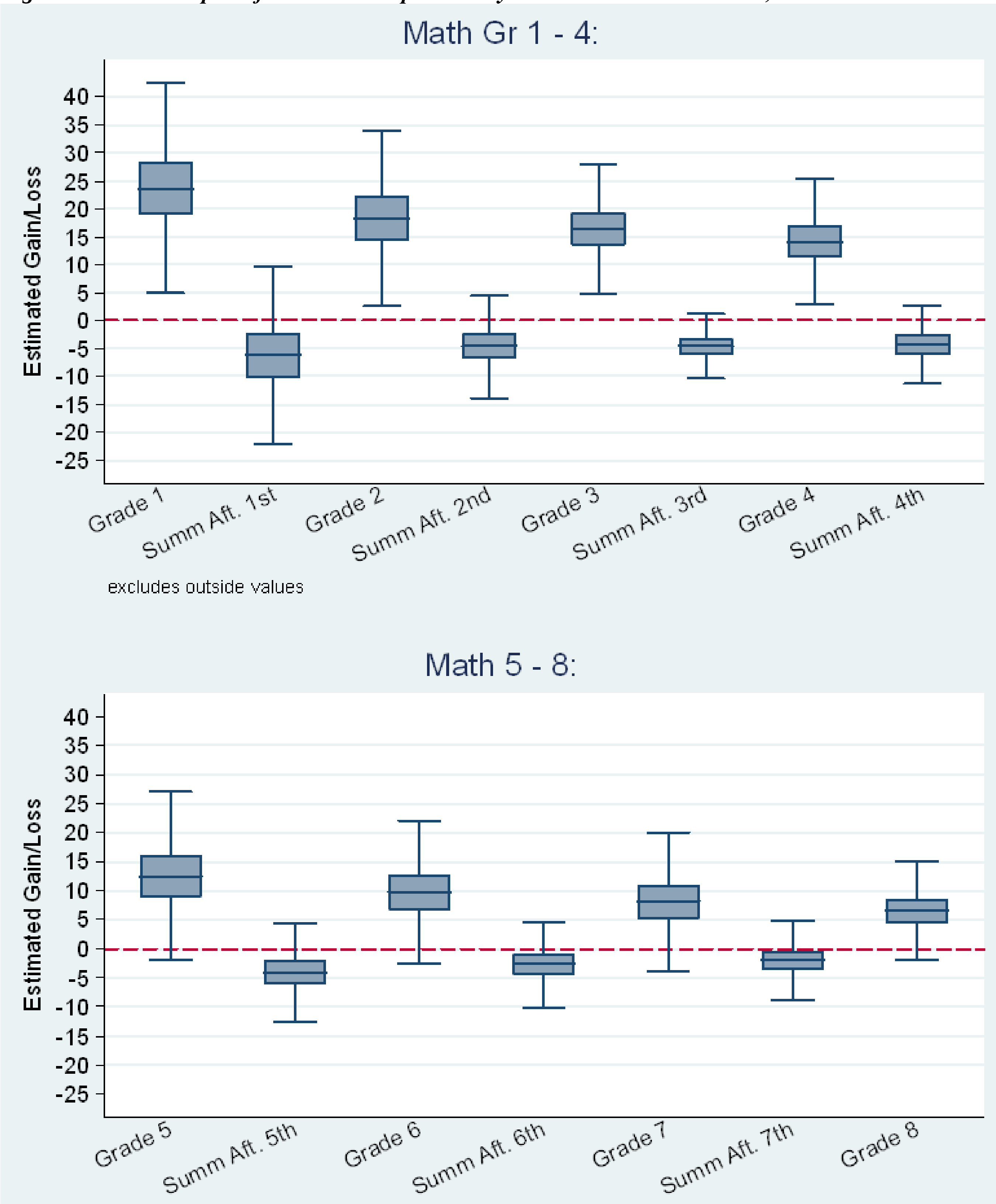


Figure 6. Math and ELA: Assume Equal Learning in School, Three Levels of Summer Gains/Losses

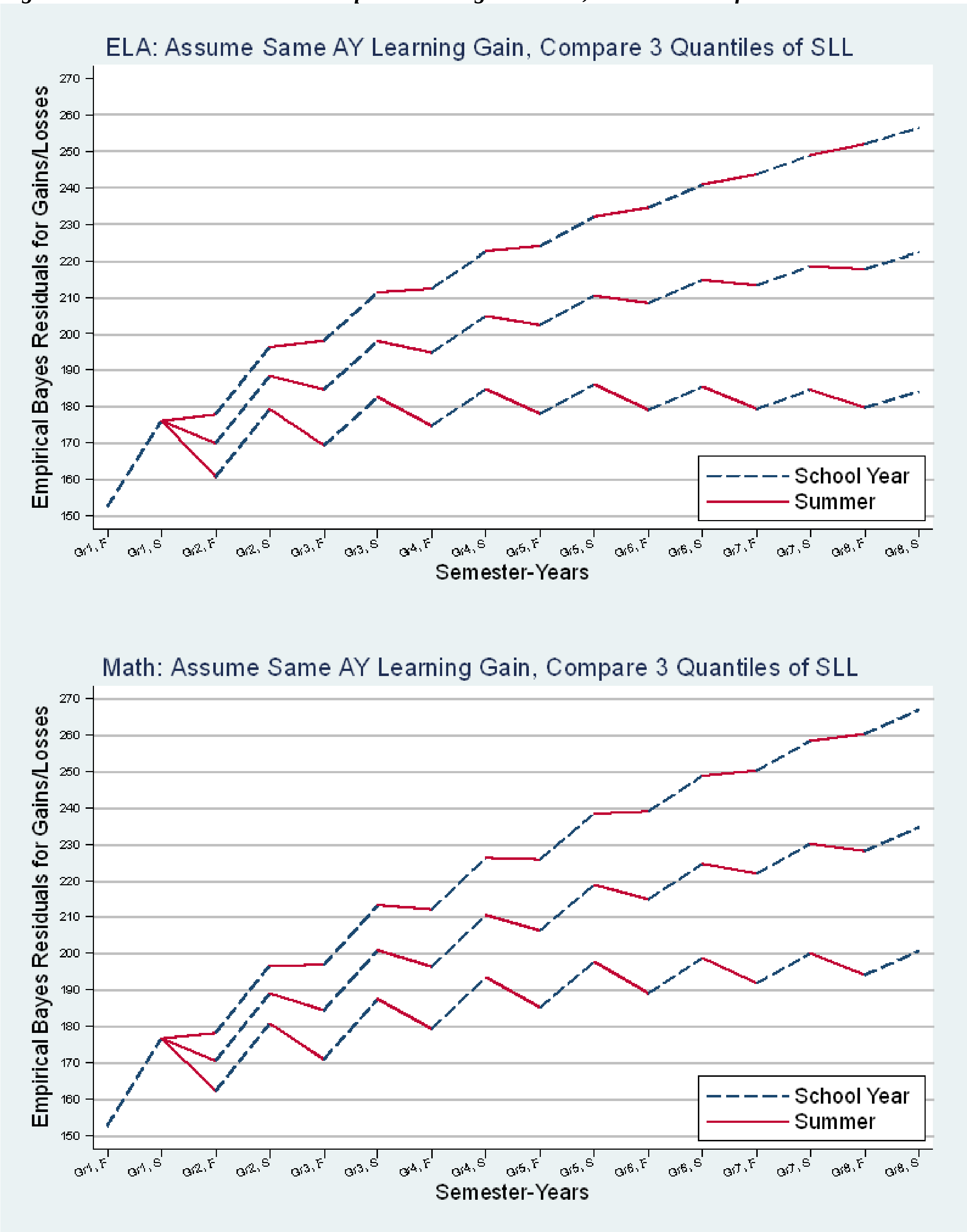
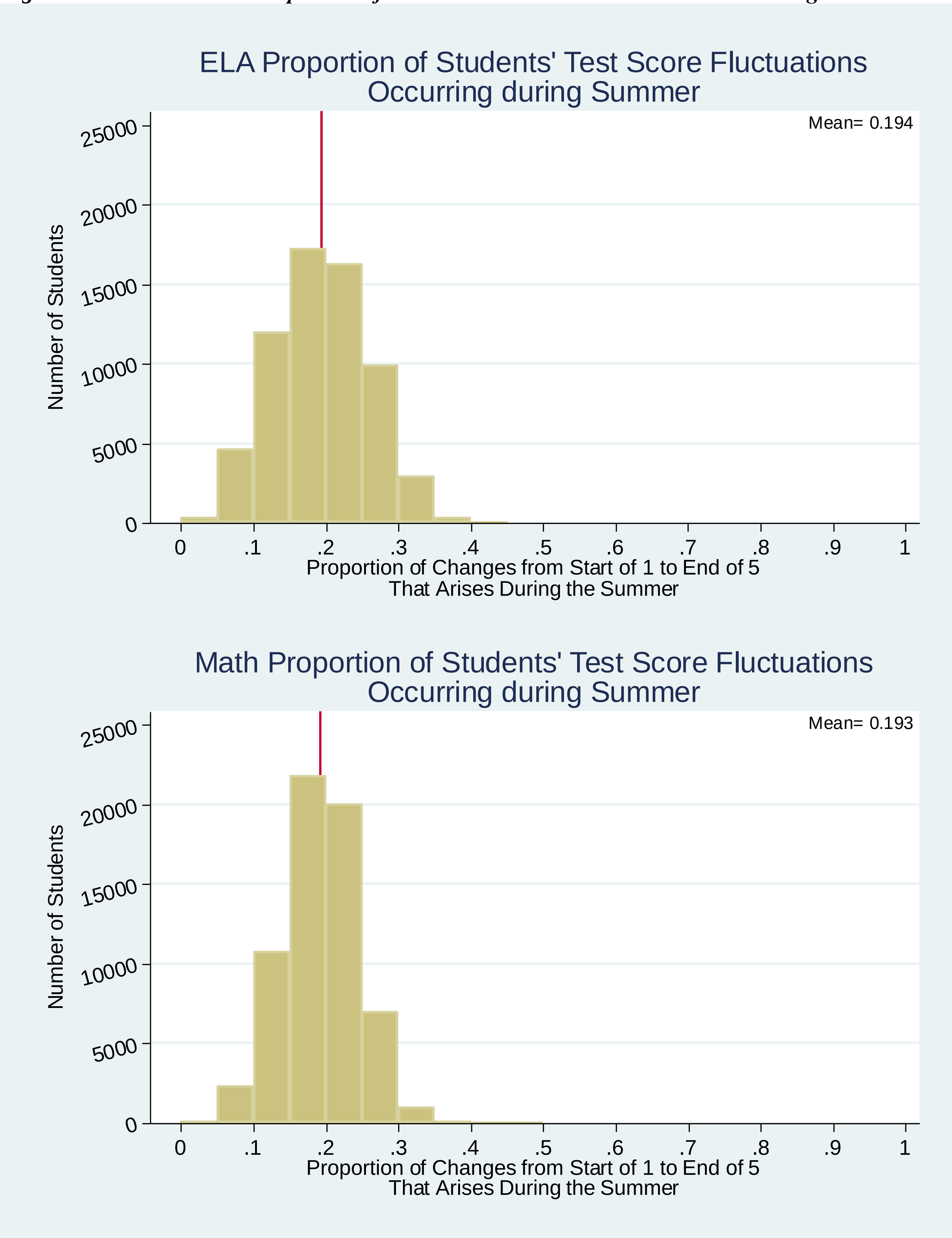


Figure 7. ELA and Math: Proportion of Students' Test Score Fluctuations Occurring in Summers



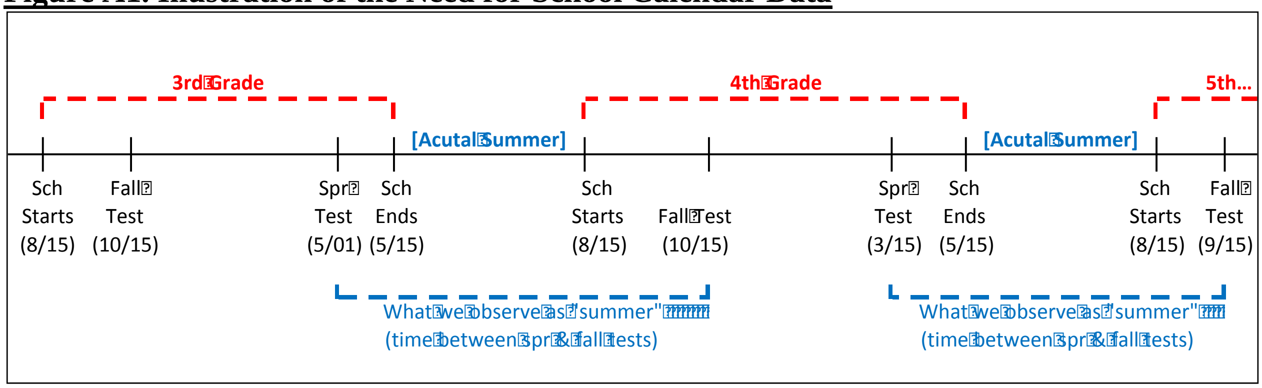
Appendix A. School Calendar Dates and Projected RIT Scores

Approach to Calendar Data Collection

One unique aspect of the current project was to collect, clean, and incorporate a new source of information about school years into both the current analyses, as well as to share the information back with our research partner NWEA to improve their own internal analyses. We collected longitudinal information on school calendars at the district level for all districts in a set of eleven states that have the largest percentage of students with MAP scores. In fact, 44.4 percent of all student-year observations from the NWEA data come from this subset of states.

In Figure 1, we show a hypothetical timeline for a given student's test-taking from 3rd through 5th grade. The Figure illustrates that students do not take MAP tests *exactly* on the first and last day of school—in fact, students often take these tests three to six weeks before/after the school year starts or ends. As a result, some of the time between the spring and fall administrations of the test is actually spent in school. However, we do not observe school-year start or end dates, leaving us with a distorted sense of how long students spend without the structure of the school year—the very time when we suspect learning rates may slow. Without knowledge of school-year calendars, we would misattribute some of the learning that takes place during the school year to the summer period, potentially masking some of the actual variation in the summer period. We therefore obtained the school-year calendar information through original data collection.

<u>Figure A1. Illustration of the Need for School Calendar Data</u>



The scope of this data collection task varied considerably, and it depended largely on whether each state has adopted a statewide policy on school year start- and end-dates, or whether state departments of education kept this information in existing data files. For example, the process for South Carolina was relatively simple because, beginning in August of 2007, South Carolina adopted new statewide legislation that specified consistent school start and end dates. We have found online a document that reported each of South Carolina school districts' calendars from 2010-11 through 2015-16. We examined the extent to which school districts actually used the uniform start and end dates mandated by the legislation (district level calendars are no longer available prior to 2010-11). In the years of district-level calendar that we have, it appears that the vast majority of South Carolina districts uses the same school year start and end dates that is described in the legislation: School typically starts on the third Monday of August, and the last day of school falls on the first Thursday of June.

In the eleven other states in which we conducted data collection, there is no statewide legislation that specifies district start and end dates. To gather the data in other states and years, we worked with a team of undergraduate and graduate student research assistants in efforts to collect complete records on school district calendars across our twelve-state sample. The first step was to exploit all online resources to find existing records from state- and district-level education departments. We also used an internet archive website (https://archive.org/web/) to search for this information that had potentially been archived in prior years. In cases where such documentation could not be found, research assistants also examined news sources archived online that document district-wide school calendars. We found that newspapers often run stories about the school year timeline. Finally, once all indirect methods of obtaining school calendar records have been exhausted, research assistants contacted appropriate district or state personnel directly to request the information.

Altogether, we proposed collecting school year start and end dates in 3,119 unique districts across eleven states and nine school years, for a total of about 28,000 district-years. We collected 23,223 school year start dates and 20,807 school year end dates. We therefore found about 77 percent of the district-year calendar dates we sought to find. In Table 1, we present the percentage of districts in each state and in each year for which we have collected school year start dates. In green we highlight cells that have over 90 percentage coverage, and in red we highlight cells that have less than 50 percent coverage.

Table A1. District Coverage (Percentage), by State and Year

State	2008	2009	2010	2011	2012	2013	2014	2015	2016
CO	23.6	23.6	23.6	25.1	24.6	25.6	26.1	86.2	77.6
IA	93.1	96.4	96.4	96.4	96.4	96.4	95.8	93.6	93.1
IN	99.1	99.1	99.1	99.2	99.2	98.7	98.7	98.7	98.7
KS	95.6	95.9	96.2	96.5	96.3	96.0	96.3	96.3	94.9
KY	10.9	9.8	100.0	100.0	100.0	100.0	99.4	99.4	99.4
MN	99.4	99.4	99.6	99.6	99.6	99.6	98.5	97.4	54.0
NH	97.7	97.7	97.7	97.7	97.7	97.7	97.2	97.2	96.7
SC	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
WA	31.9	43.4	61.1	65.5	65.2	63.9	75.7	91.2	96.3
WI	99.3	98.2	98.6	99.5	99.6	99.6	99.6	99.6	0.0
WV	100.0	100.0	100.0	98.2	100.0	100.0	100.0	100.0	0.0

In later years, NWEA also began to collect some school-year start and end dates. We combined our original data collection with theirs. Across the entire NWEA dataset in all states and years, these efforts allowed us to collect actual calendar start/end dates for 50.3 percent of the observed school-years. We refer to these as the "actual calendar dates", because we also opt to extrapolate calendar dates for all districts in which they are missing.

Using Actual Calendar Dates to Extrapolate Missing Calendar Dates

In order to project scores for students in districts for which we were unable to recover actual calendar dates, we chose to impute approximate school calendar dates under the basic assumption that, while there is some variation in when public school districts start and end the school-year, it is not large. For example, in the subset of districts for which NWEA collected *school* level calendar start dates, we observe that the standard deviation of start dates across schools in the same district and same year is 8.2 days (8.1 days for end dates). Looking across all the districts in a given state in the same year, the standard deviation of start dates is 6.3 days (8.2 for end dates).

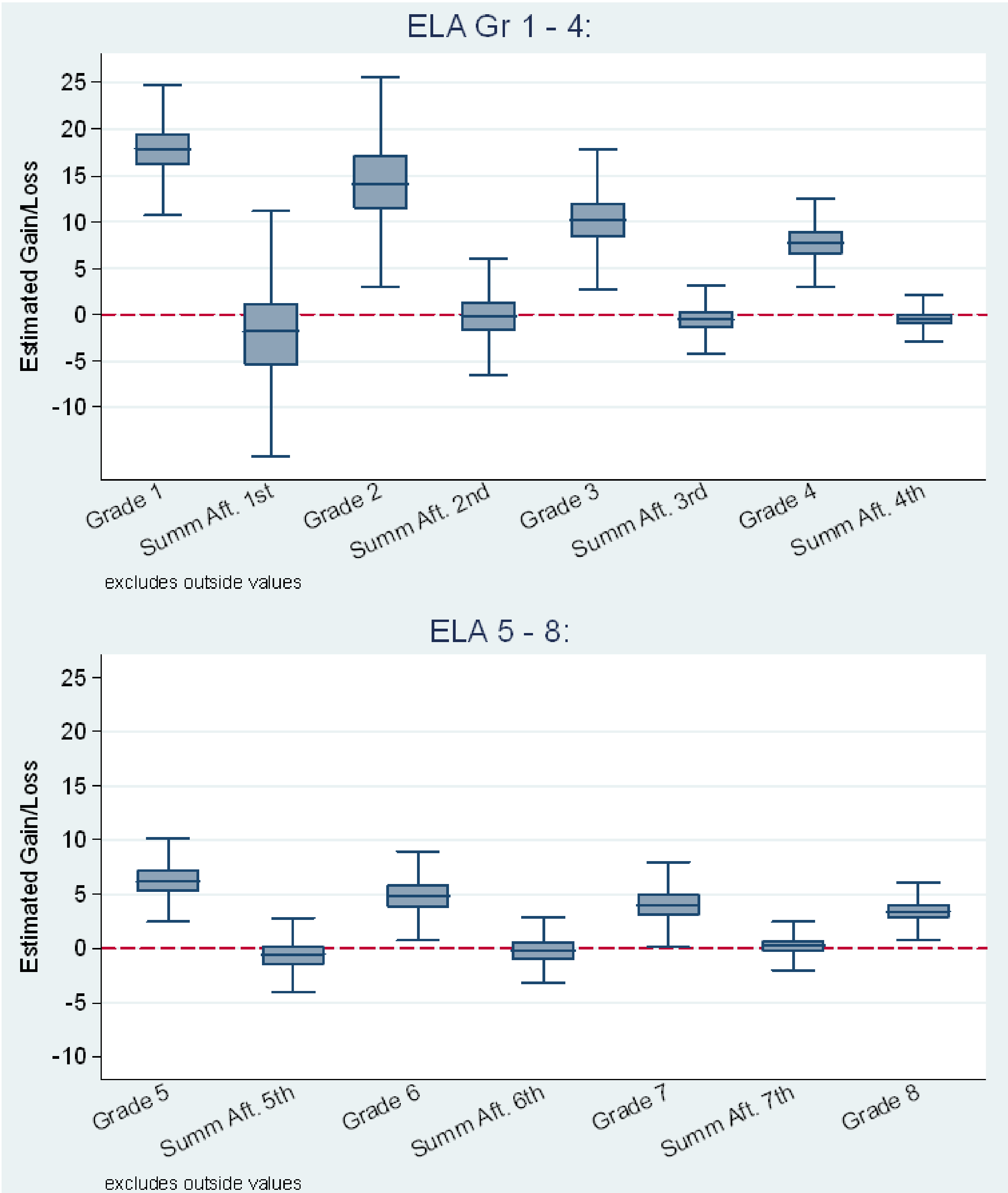
We therefore extrapolated dates privileging these decision rules in the following order: (1) If we have actual school calendar data, use that. (2) For schools in a district-year with some school calendar data, use the mean of the start/end dates in the district-year. (3) For a district that has calendar data in some years but not others, use the district's own mean start/end dates across years. (4) For districts still missing start/end calendar dates, use the state's mean dates in the given year. (5) For districts in states that have no calendar data in a given year, use the state's mean calendar dates across all years. Because we had calendar data in at least one year for each state, this covered all observations in the dataset.

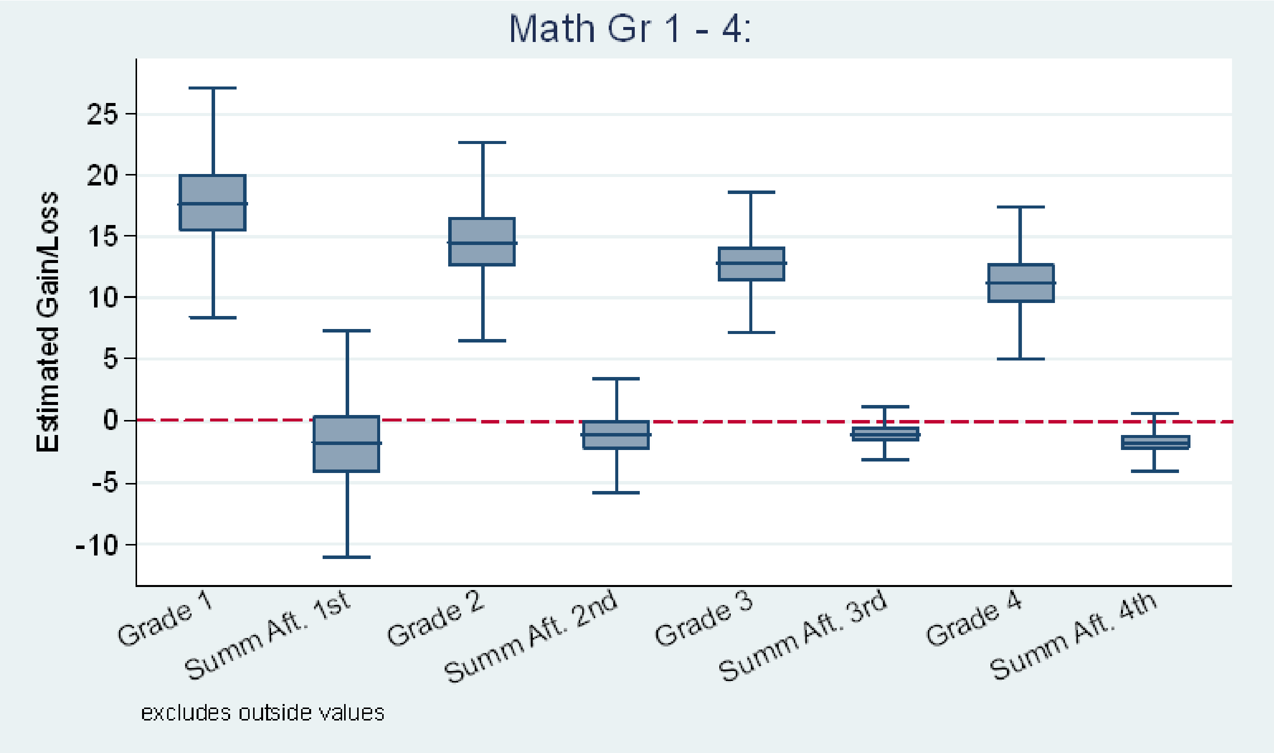
Projecting RIT Scores to First and Last Day of School

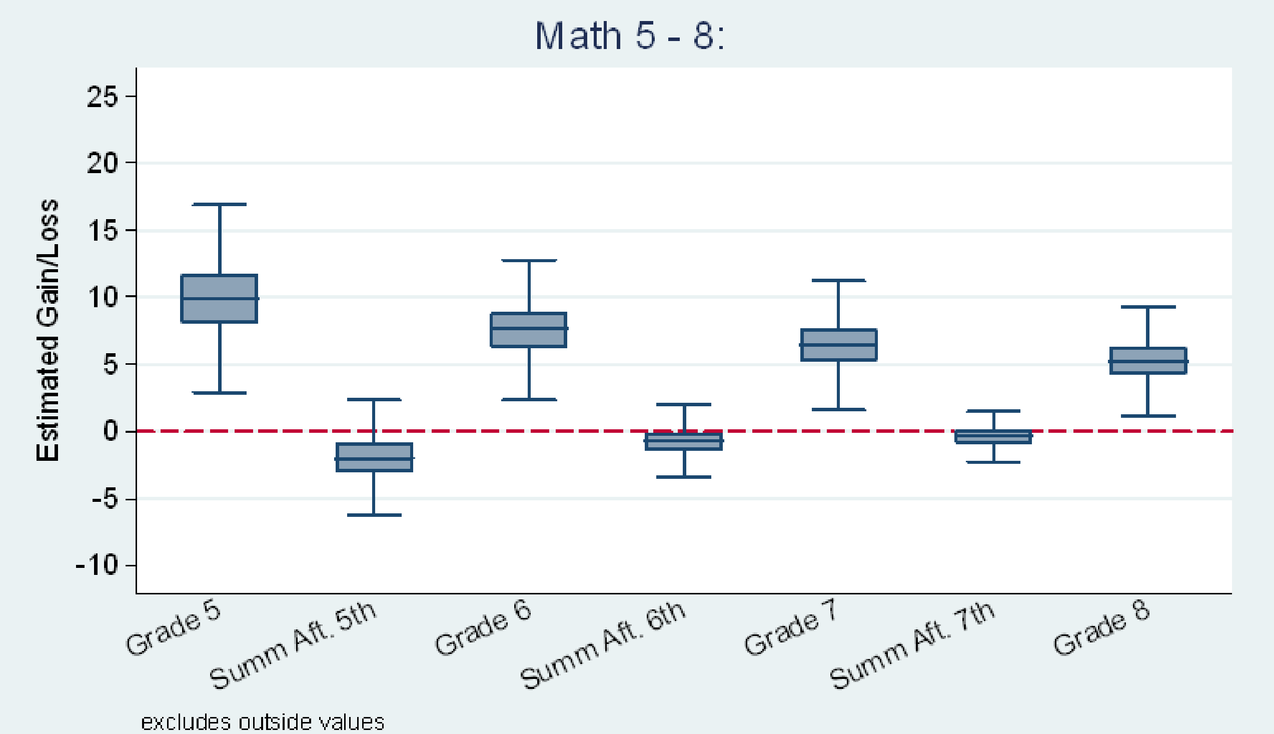
We leverage the calendar data described above to project scores for individual students to what they might have been on the first and last day of school. To do so, we calculate the average daily learning rate between each student's fall and spring NWEA test administrations by dividing the change in score by the number of days between the two tests (Quinn, 2014). Extant research finds that students' within school-year achievement growth is approximately linear (Fitzpatrick, Grissmer, & Hastedt, 2011). We then calculate both the number of school days between the start of the school year and each student's fall NWEA test, as well as the number of days of school between each student's spring NWEA and the end of the school year. On average, students take the fall test about 26 days after the first day of school, and they take the spring test 39 days before the last day of school.

To project scores to the start of the school year, we subtract from the student's observed fall score his or her individual daily learning rate multiplied by the number of days between the first day of school and the date of the test. We follow the same procedure for projecting scores to the last day of the school year. The correlation between fall observed and projected scores in ELA is 0.996, with an RMSE of 2.3 points. The correlation between spring observed and projected scores in ELA is 0.992, with an RMSE of 2.8 points.

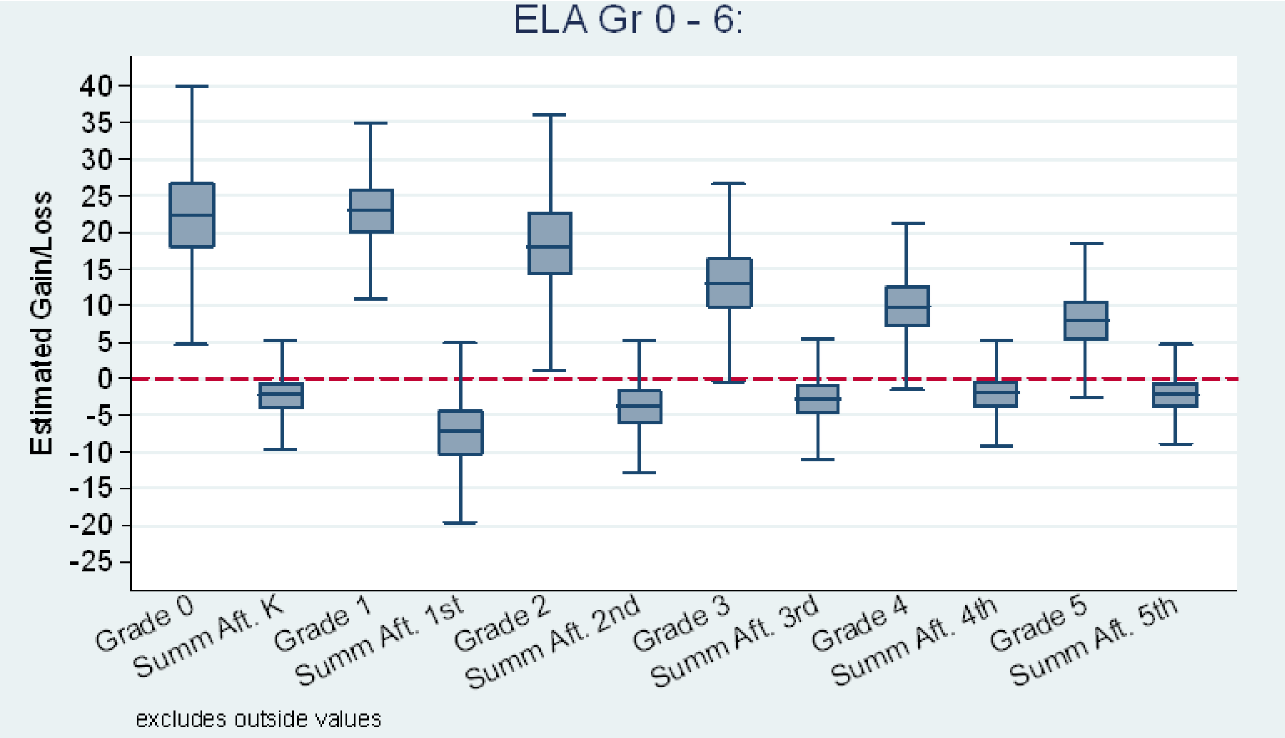
Appendix A: Replicate Figure 3 and 4 using Observed Scores

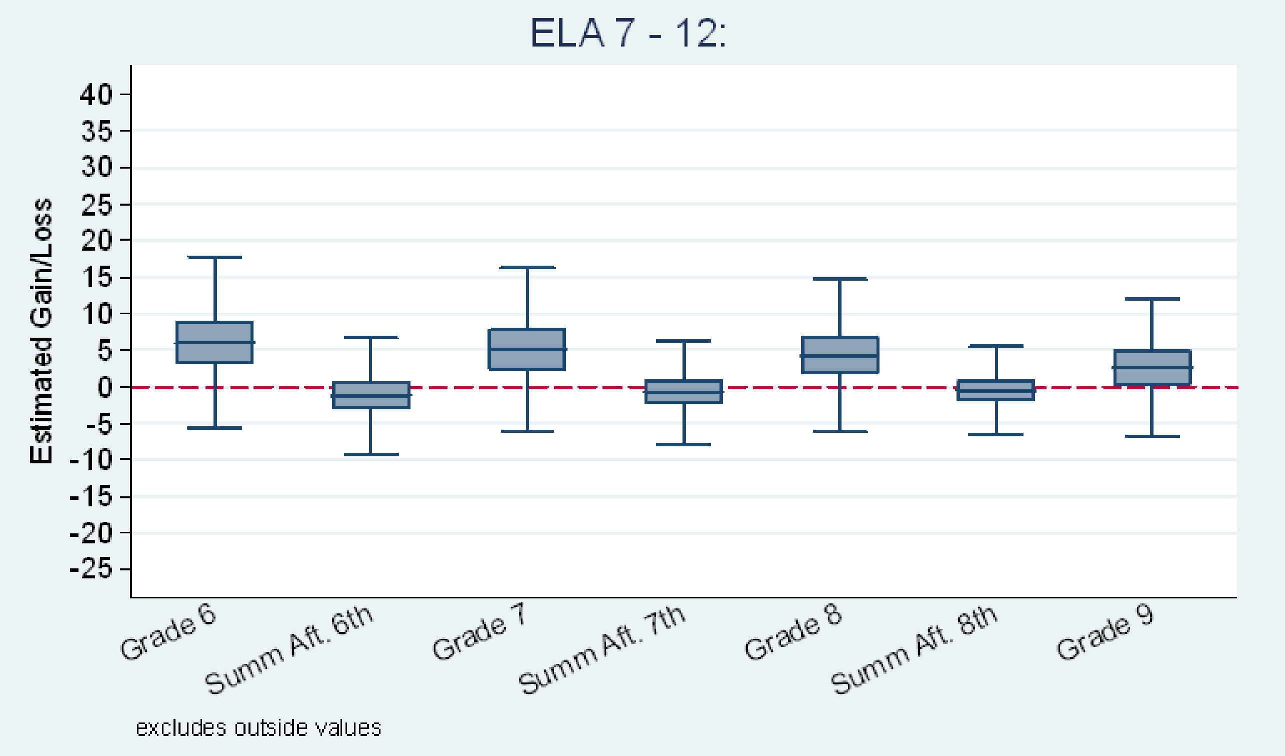


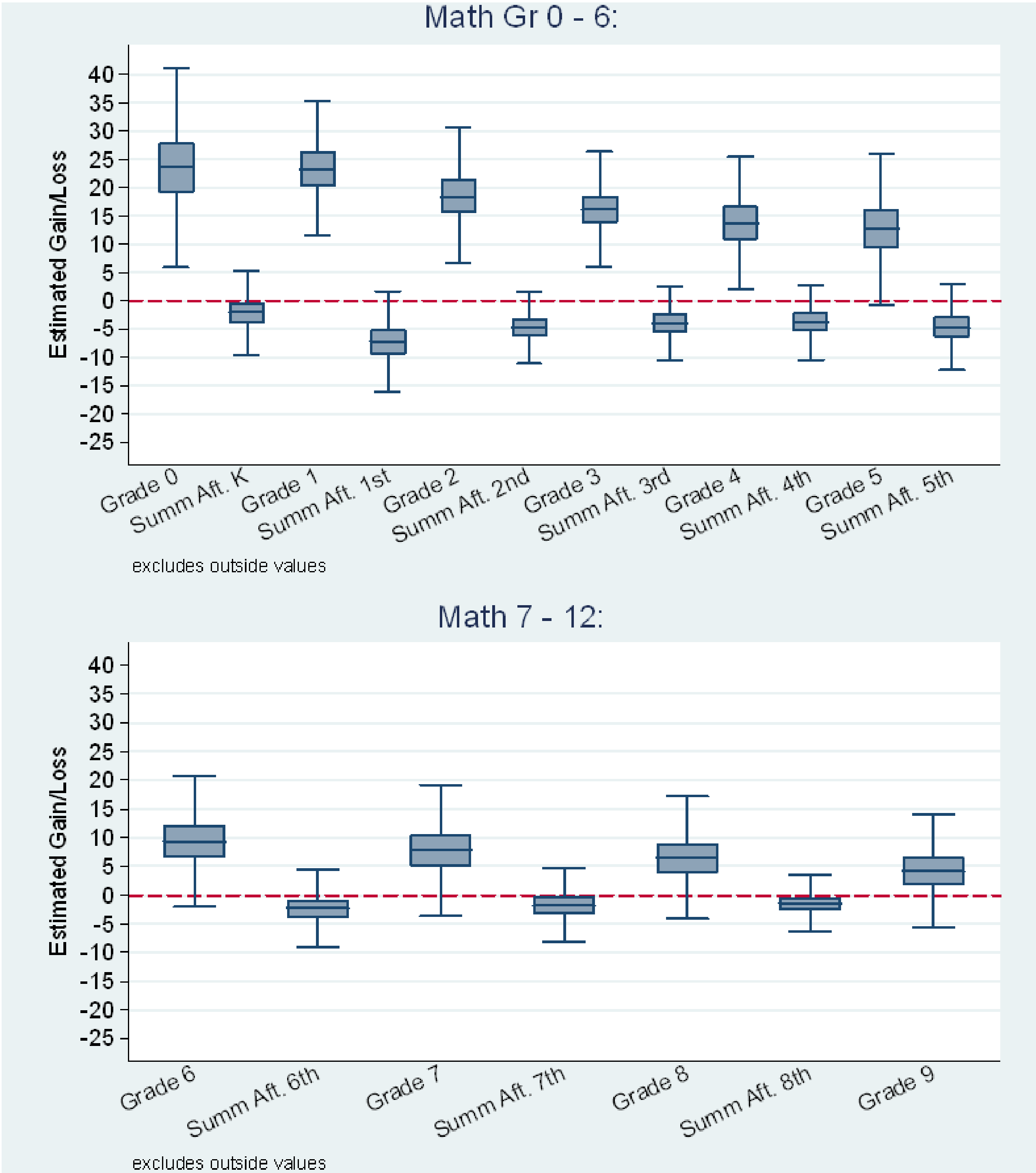

Math Gr 1 - 4:
Estimated Gain/Loss
25
20
15
10
5
0
-5
-10
Grade 1
Summ Aft. 1st
Grade 2
Summ Aft. 2nd
Grade 3
Summ Aft. 3rd
Grade 4
Summ Aft. 4th
excludes outside values


Math 5 - 8:
Estimated Gain/Loss
25
20
15
10
5
0
-5
-10
Grade 5
Summ Aft. 5th
Grade 6
Summ Aft. 6th
Grade 7
Summ Aft. 7th
Grade 8
excludes outside values

Appendix B. Replicate Figure 4 & 5 with 3-Grade Increment







Math Gr 0 - 6:
Estimated Gain/Loss
40
35
30
25
20
15
10
5
0
-5
-10
-15
-20
-25
Grade 0
Summ Aft. K
Grade 1
Summ Aft. 1st
Grade 2
Summ Aft. 2nd
Grade 3
Summ Aft. 3rd
Grade 4
Summ Aft. 4th
Grade 5
Summ Aft. 5th
excludes outside values
Math 7 - 12:
Estimated Gain/Loss
40
35
30
25
20
15
10
5
0
-5
-10
-15
-20
-25
Grade 6
Summ Aft. 6th
Grade 7
Summ Aft. 7th
Grade 8
Summ Aft. 8th
Grade 9
excludes outside values

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*(2), 167.

Allinder, R. M., Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1992). Effects of summer break on math and spelling performance as a function of grade level. *The Elementary School Journal*, 451-460.

Benson, J., & Borman, G. (2010). Family, Neighborhood, and School Settings across Seasons: When Do Socioeconomic Context and Racial Composition Matter for the Reading Achievement Growth of Young Children? *Teachers College Record, 112*(5), 1338-1390.

Borman, G. D., Benson, J., & Overman, L. T. (2005). Families, schools, and summer learning. *The Elementary School Journal, 106*(2), 131-150.

Briggs, D. C. (2013). Measuring growth with vertical scales. *Journal of Educational Measurement, 50*(2), 204-226.

Briggs, D. C., & Weeks, J. P. (2009). The impact of vertical scaling decisions on growth interpretations. *Educational Measurement: Issues and Practice, 28*(4), 3-14.

Burkam, D. T., Ready, D. D., Lee, V. E., & LoGerfo, L. F. (2004). Social-class differences in summer learning between kindergarten and first grade: Model specification and estimation. *Sociology of Education, 77*(1), 1-31.

Cann, D., Karakaplan, M., Lubke, M., & Rowland, C. (2015). *New Mexico StartSmart K-3 Plus Validation Study: Evaluator's Report* Retrieved from Presented at APPAM Conference 2014.: http://ccpi.unm.edu/sites/default/files/publications/EvaluatorReport.pdf

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227.

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*(5), 613.

Downey, D. B., von Hippel, P. T., & Hughes, M. (2008). Are "failing" schools really failing? Using seasonal comparison to evaluate school effectiveness. *Sociology of Education, 81*(3), 242-270.

Entwisle, D. R., & Alexander, K. (1990). Beginning school math competence: Minority and majority comparisons. *Child Development, 61*(2), 454-471.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review*, 72-84.

Entwisle, D. R., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171.

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.

Gershenson, S. (2013). Do summer time-use gaps vary by socioeconomic status? . *American Educational Research Journal, Forthcoming, .*

Gilkerson, J., & Richards, J. A. (2009). The power of talk. *Impact of adult talk, conversational turns and TV during the critical 0-4 years of child development: Boulder, CO: LENA Foundation.*

Graves, J. (2011). Effects of year-round schooling on disadvantaged students and the distribution of standardized test performance. *Economics of Education Review, 30*(6), 1281-1305.

Halle, T., Forry, N., Hair, E., Perper, K., Wandner, L., Wessel, J., & Vick, J. (2009). Disparities in early learning and development: lessons from the Early Childhood Longitudinal Study–Birth Cohort (ECLS-B). *Washington, DC: Child Trends*.

Heyns, B. (1978). *Summer learning and the effects of schooling*: Academic Press New York.

Kaushal, N., Magnuson, K., & Waldfogel, J. (2011). *How is family income related to investments in children's learning?* : Russell Sage Foundation.

Kim, J. S., & Quinn, D. M. (2013). The Effects of Summer Reading on Low-Income Children's Literacy Achievement From Kindergarten to Grade 8 A Meta-Analysis of Classroom and Home Interventions. *Review of Educational Research*.

Kornrich, S., & Furstenberg, F. (2013). Investing in Children: Changes in Parental Spending on Children, 1972–2007. *Demography, 50*(1), 1-23.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*: ERIC.

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal, 41*(1), 115-157.

McCombs, J. S., Augustine, C., Schwartz, H., Bodilly, S., McInnis, B., Lichter, D., & Cross, A. B. (2012). Making Summer Count: How Summer Programs Can Boost Children's Learning. *Education Digest: Essential Readings Condensed for Quick Review, 77*(6), 47-52.

McCombs, J. S., Pane, J. F., Augustine, C. H., Schwartz, H. L., Martorell, P., & Zakaras, L. (2015). First Outcomes from the National Summer Learning Study.

McEachin, A., & Atteberry, A. (2017). The impact of summer learning loss on measures of school performance. *Education Finance and Policy, 12*(4), 468-491.

McMullen, S. C., & Rouse, K. E. (2012). The Impact of Year-Round Schooling on Academic Achievement: Evidence from Mandatory School Calendar Conversions. *American Economic Journal: Economic Policy, 4*(4), 230-252.

Quinn, D. M. (2014). Black-White Summer Learning Gaps: Interpreting the Variability of Estimates Across Representations. *Educational Evaluation and Policy Analysis*. doi:10.3102/0162373714534522

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

Reardon, S. F. (2014). *Education. In State of the Union: The Poverty and Inequality Report 2014*. Retrieved from

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*: National Academies Press.

Skibbe, L. E., Grimm, K. J., Bowles, R. P., & Morrison, F. J. (2012). Literacy growth in the academic year versus summer from preschool through second grade: Differential effects of schooling across four skills. *Scientific Studies of Reading, 16*(2), 141-165.