IDENTIFICATION OF SUMMER SCHOOL EFFECTS BY COMPARING THE IN- AND OUT-OF-SCHOOL GROWTH RATES OF STRUGGLING EARLY READERS

ABSTRACT

A one-group repeated-treatment design was used to examine the academic year and summer oral reading fluency outcomes for students attending a district-sponsored summer literacy program ($N = 250$). Piecewise growth models applied to longitudinal data obtained during the first and second grade and over the course of the intervening summer revealed that oral reading fluency increased during each period of schooling, with the most rapid increase occurring during the intensive summer school intervention period. The gains in reading fluency observed during periods of schooling contrasted with periods of stagnation or loss when students were not in school during each of two summer breaks. The observed pattern of learning suggests that for the struggling readers we studied, schooling "mattered" regardless of when in the calendar year it was experienced. Challenges and opportunities associated with evaluating summer program performance are discussed.

Keith Zvoch
Joseph J. Stevens
UNIVERSITY OF OREGON

THE pattern of achievement growth established in the analysis of time-series data obtained on individual students reflects a clear upward learning trajectory during the academic year, but a slowing or loss of learning over the summer vacation period (Alexander, Entwisle, & Olson, 2001; Downey, von Hippel, & Broh, 2004; Downey, von Hippel, & Hughes, 2008; Raudenbush, 2004; Skibbe, Grimm, Bowles, & Morrison, 2012). Learning-rate differences across in- and out-of-school periods serve as clear evidence of the magnitude of schooling impacts

THE ELEMENTARY SCHOOL JOURNAL VOLUME 115, NUMBER 3

on student achievement growth, but provide little solace for school personnel held accountable for either the status-based or calendar year performance of students (von Hippel, 2009). Accountability for the learning that occurs during the academic year and summer recess period has stimulated interest in supplemental instructional initiatives delivered during the summer months (McCombs et al., 2011). For school personnel, strategically targeted summer instruction is increasingly viewed as a cost-effective mechanism for keeping students most at risk of summer learning losses scholastically engaged (Fairchild, Smink, & Stewart, 2009; McCombs et al., 2011; Smink, 2011). Yet, the need-based delivery of summer instruction often presents an evaluative challenge as invitations to school-based summer learning programs are distributed nonrandomly by design. The need-based assignment practices typically employed by school personnel result in either no comparison group or nonequivalent group designs that make the separation of summer instructional effects from the background and motivational characteristics of students and families difficult. The inferential challenge posed by the naturally occurring but weak quasi-experimental designs that arise in conjunction with the nonrandom delivery of summer school motivates the following investigation. Specifically, the purpose of this article is to highlight a methodological/analytic approach that can be used to evaluate the impact of summer instruction when equivalent comparison or control groups are not possible. A second intent is to fully utilize a rich source of longitudinal data to clearly and more accurately demonstrate the in- and out-of-school learning trajectories of struggling early readers and describe associated implications for the evaluation of summer program performance.

In the past decade, passage of the federal No Child Left Behind (NCLB, 2002) and the Individuals with Disabilities Education Improvement (IDEA, 2004) Acts has set in motion a series of state, district, and school-based administrative and organizational changes that have transformed the manner in which instruction is delivered and student and school performance is measured and evaluated. With enactment of NCLB and IDEA, K–12 stakeholders have attempted to identify and implement programs and practices to raise proficiency levels and narrow the achievement gap between advantaged and disadvantaged student groups. The push to engender positive academic outcomes for all students has led to an increase in the use of initiatives that restructure or extend the academic day (e.g., full-day kindergarten, after school programs), the academic calendar (e.g., year-round schooling), and the daily delivery of targeted supplemental instructional support for students most at risk of negative academic outcomes (see, for reviews, Cooper, Batts Allen, Patall, & Dent, 2010; Cooper, Valentine, Charlton, & Melson, 2003; Gersten et al., 2008; Lauer et al., 2006; McCombs et al., 2011; Patall, Cooper, & Batts Allen, 2010).

For many students, access to a longer and/or more effective instructional day aids in offsetting home- and community-based resource challenges. However, a factor that limits the extent to which schools and school-based initiatives can directly promote student achievement success is the time that students spend outside the scholastic environment. During the academic year, weekend and quarter breaks provide students a respite from daily instruction. The absence of formal school-based instruction is considerably greater in the summer. A traditional academic calendar offers students 3 months of "vacation" from schooling. Long summer breaks are problematic for schools struggling to meet NCLB proficiency objectives as summer learning slows (relative to school year learning) for students of advantage and flattens or declines for disadvantaged students (Alexander et al., 2001; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Heyns, 1978, 1987). On average, students lose approximately 1 month of grade-equivalent skills over the summer (Cooper et al., 1996). However, summer achievement outcomes are not uniform across content area and poverty level as mathematics losses are greater than reading losses and more advantaged students tend to be less affected by the summer break than their more disadvantaged peers (Benson & Borman, 2010; Burkam, Ready, Lee, & LoGerfo 2004; Cooper et al., 1996). More problematically, the differential rates of summer learning tend to cumulate over time and contribute to the widening achievement gap between advantaged and disadvantaged students (Alexander et al., 2001; Alexander, Entwisle, & Olson, 2007).

To combat the differential and cumulative impact of summer learning losses, school districts have increasingly turned to the use of supplemental summer instruction. A recent estimate indicates that approximately 14.3 million school age children (25% of the school age population) attended a summer learning program during the summer of 2008 (America After 3PM, 2010). Participation rates were higher among low-income and ethnic minority students, but the unmet demand for summer educational programs was also highest among these special population groups (America After 3PM, 2010). Although widely viewed as a positive development in the effort to prevent the loss of academic year gains, the evidence linking summer school with summer learning has historically been limited by weak evaluation designs (Cooper, Charlton, Valentine, & Muhlenbruck, 2000). Yet, in their rigorous review of studies, summer school programs were estimated to have a quarter of a standard deviation effect on the achievement outcomes of participants (Cooper et al., 2000). More recent studies that have been based on randomized (Borman & Dowling, 2006; Borman, Goetz, & Dowling, 2009; Schacter & Jo, 2005; Zvoch & Stevens, 2013) or regression discontinuity (Jacob & Lefgren, 2004; Matsudaira, 2008; Zvoch & Stevens, 2011) designs also provide evidence of positive summer school effects. In these studies, summer school students tended to either outgain their comparison-group peers during the summer vacation period or produce higher scores at the beginning of the following academic year on various reading and mathematics measures.

In combination, the literature on summer learning rates and supplementary summer instruction highlights the summer learning loss phenomenon, documents various programmatic attempts to offset the summer slide, and elucidates a range of methodological issues that are germane to the validity of inferences drawn from studies designed to estimate summer school effects (Cooper et al., 1996, 2000; McCombs et al., 2011). The literature also identifies a measurement artifact that is typically operative in summer learning investigations as a portion of academic year learning gains are captured in traditional summer learning estimates. More specifically, as the assessments on which estimates of summer learning are based are not given on the last day of classes in one academic year and the first day of classes in the subsequent year, summer learning rates are partially contaminated by the school-based learning that takes place after the final assessment in the spring and before the initial assessment in the fall (Burkam et al., 2004; Cooper et al., 1996; Heyns, 1978, 1987). Consequently, summer learning losses are underestimated to the extent that achievement gains accrue during the lag between the final assessment at the end of one academic year (i.e., the pretest) and before the initial assessment in the fall of the subsequent year (i.e., the posttest). In an attempt to more accurately estimate summer changes in achievement, some authors have used "clock" variables to take account of intervals from the last assessment in spring through summer break to the next assessment in the fall (Burkam et al., 2004; Downey et al., 2004, 2008, McCoach, O'Connell, Reis, & Levitt, 2006). Yet, in the majority of cases, no adjustment for the instruction that occurs during the assessment lags is performed, and, to our knowledge, there are currently no summer learning/summer school studies that have been conducted using test score data directly collected during the summer.

From a methodological perspective, the evaluative complexities that arise in conjunction with supplemental summer instruction have been an unwelcome barrier to identifying whether and to what extent summer school interventions effectively impact the summer learning of at-risk students. However, when the time periods in which students are exposed to school-based instruction are recorded and achievement outcomes are contiguously measured, researchers can draw on the power of the associated repeated-treatment (RT) design to increase the validity of inferences regarding the relative treatment effect of supplemental summer instruction as well as more accurately identify the trajectory of learning that occurs across the calendar year. An RT design can usefully be applied in a range of settings where a treatment (e.g., school-based instruction) is repeatedly introduced and removed and a comparison group is not available. RT designs yield strong inferences when performance trends contiguously align with the initiation, removal, and reintroduction of treatment. Strength of inference is further enhanced as multiple replications are conducted and as an abrupt reversal in the outcome can be demonstrated at the points at which the intervention is introduced and removed (Shadish, Cook, & Campbell, 2002).

The current study utilizes an RT design to estimate and compare the in- and out-of-school learning rates of a sample of at-risk readers. With the repeated assessment of individual students and knowledge of the timing of instructional exposures over 2 academic years and the intervening summer, estimation of changes in student learning associated with the delivery and removal of instruction served to illustrate the impact of an academically rigorous summer school program. The intervention context is that of traditional academic year schooling and a 5-week summer literacy intervention delivered to early struggling readers. The presence of naturally occurring academic calendar breaks prior to and following the summer intervention and a strategically implemented assessment schedule provided the means by which period-specific learning rates were estimated and evaluated. In all, the availability of nine observations spanning the course of the first- and second-grade school years and the summer intervention period enabled estimation and comparison of (1) academic year and summer school growth rates and (2) the change in performance across each out-of-school period. Hypotheses were that student learning would be positive during each academic year and the summer instructional intervention, but between school enrollments, the rate of learning would flatten or decline.

Method

Data Source

The present study was conducted as part of a larger evaluation of a summer literacy program implemented in a moderate-sized Pacific Northwest school district. Each year, the district serves approximately 6,000 K–12 students. In recent years, the student body has identified as 75% White, 14% Latino, 3% African American, 3% Asian American, 3% Native American, and 2% other. Approximately 44% of district students receive a free- or reduced-price lunch, and 3% of district students are classified as English language learners.

Analytic Sample

Data obtained from students in three successive first-grade student cohorts (i.e., 2008–2009, 2009–2010, and 2010–2011) were analyzed ($N = 250$). For each cohort, students' early literacy skill development was tracked across the first and second grades and over the intervening summer. The cohorts (cohort 1, $n = 86$; cohort 2, $n = 76$; cohort 3, $n = 88$) consisted of the subset of district students who were invited to and participated in summer school. Summer school participants were retained in the study even if they missed one or more academic year or summer school assessment occasions. The district offered need-based summer school placements to students falling below or in an interval slightly above benchmark performance on a spring interim assessment. The distribution of need-based placements resulted in an analytic sample that was more disadvantaged than the general student body. Specifically, 72% ($n = 181$) of summer school participants received a free- or reduced-price lunch during the academic year, 29% ($n = 72$) identified with an ethnic minority group, 12% ($n = 29$) were classified as English language learners, and 16% ($n = 41$) were offered special education services. The analytic sample also consisted of 50% girls ($n = 124$). During the academic year, students were enrolled in one of seven district elementary schools.

Summer School Program

A 5-week supplemental literacy program for struggling readers was conducted during the middle of the 3-month summer vacation period. Each summer, the academically intensive program began in mid-July and ended in mid-August, starting approximately 4 weeks after the end of first grade and ending approximately 4 weeks prior to the start of second grade. The summer program was designed to help struggling early readers gain the foundational skills requisite to meet reading proficiency targets and otherwise shrink the literacy gap with their higher-performing peers. During the summer intervention, instruction occurred 3.5 hours/day, four mornings per week in small class size environments housed in one central school site. Each instructional day students received whole-group instruction, intensive direct instruction in small groups, and a 15-minute recess followed by a second period of small-group direct instruction. Small-group leaders provided direct instruction in phonemic awareness, the alphabetic principle, and fluency and comprehension in each of the two daily 45-minute sessions. Instructional content and practices were aligned with the fundamental "big ideas" and best practices that have been shown to facilitate early childhood literacy development (National Reading Panel, 2000). Key instructional components were teacher skill modeling, frequent opportunity for student practice, and individually tailored instruction and feedback (Zvoch & Stevens, 2011).

Outcome Measure

Oral reading fluency (ORF) scores were used to measure changes in literacy skill development. Oral reading fluency refers to the automatic, effortless, and accurate reading of connected text and serves as a key literacy outcome in the early grades. ORF is considered a higher-order skill that integrates lower-level processes (e.g., letter naming, phonological awareness, decoding) critical to word recognition (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Slocum, Street, & Gilbert, 1995). There is strong theoretical and empirical support for the validity of scores obtained from ORF measures, including a substantial literature that demonstrates that reading fluency functions as an indicator of comprehension and overall reading achievement (e.g., Baker et al., 2008; Fuchs et al., 2001; Kim, Petscher, Schatschneider, & Foorman, 2010; Reschly, Busch, Betts, Deno, & Long, 2009). In addition, ORF scores have also been linked to reading performance outcomes on an array of year-end, high-stakes state reading tests (Crawford, Tindal, & Stieber, 2001; Good, Simmons, & Kame'enui, 2001; McGlinchey & Hixson, 2004; Nese, Park, Alonzo, & Tindal, 2011; Silberglitt, Burns, Madyun, & Lail, 2006; Wood, 2006). ORF's predictive power is thought to derive from its ability to index the efficiency of word-recognition processes. To the extent that a child is able to automatically and effortlessly identify words and read written text, attention is freed to focus on word meaning and text comprehension (Fuchs et al., 2001).

Oral reading fluency was evaluated on the basis of scores obtained from administration of the Test of Oral Reading Fluency (TORF; Children's Educational Services, 1987). The TORF is an individually administered, curriculum-based measure of fluency with connected text. The TORF was developed to identify struggling readers and allow the monitoring of student progress over time. Student performance is assessed by having students read each of three passages aloud for 1 minute. Omitted or substituted words and hesitations longer than 3 seconds are marked as errors, unless students self-correct within a 3-second timeframe. The oral reading fluency rate is based on the median words correct per minute across the three passages.

The test developer reported that test-retest reliabilities obtained from the repeated assessment of elementary students ranged from .92 to .97, and alternate-form reliabilities ranged from .89 to .94 (Tindal, Marston, & Deno, 1983). Moderate to strong concurrent relationships between TORF scores and performance on a variety of other passage fluency and reading criterion measures have been observed (Good & Jefferson, 1998; Marston, 1989). Relatively strong predictive relationships between TORF scores and later performance on the third-grade Oregon Statewide Assessment reading/language subtest have also been reported ($r = .67$ to .82; Good et al., 2004). During the academic year, the TORF is administered to first- and second-grade students during the first week of classes in mid-September, immediately following the winter break at the beginning of January, and toward the end of the academic year in mid-May.[1] In the summer, program participants completed the TORF three additional times, during the first (mid-July), third (end of July/early August), and fifth (mid-August) weeks of the summer school intervention. TORF assessments were administered over a 5-day testing window during the academic year and during a 2-day period during the summer. First- and second-grade TORF passages were designed to be of equal difficulty within and between grade level (Children's Educational Services, 1987). Forty different TORF passages were administered

Table 1. TORF Descriptive Statistics by Assessment Occasion

TORF Assessment	Mean	*SD*	*N*	Skew	Kurtosis	Min, Max Score
G1 mid-September	6.47	2.24	47	−.81	.49	0, 10
G1 beg-January	12.63	6.09	237	.64	1.17	0, 36
G1 mid-May	26.38	12.85	250	.82	.37	0, 73
SS mid-July	26.48	12.16	242	.58	.29	3, 71
SS end-July/beg-August	29.55	15.54	231	.68	.80	3, 90
SS mid-August	34.91	16.50	227	.54	.20	4, 89
G2 mid-September	32.49	16.65	234	.44	−.20	1, 83
G2 beg-January	53.65	23.13	228	.06	−.73	4, 110
G2 mid-May	77.54	26.27	224	−.25	.14	9, 157

Note.—G1 = grade 1, SS = summer school, G2 = grade 2.

over the course of the study. Descriptive statistics associated with the TORF assessment at each measurement occasion are presented in Table 1.

With respect to published reading fluency norms (Hasbrouck & Tindal, 2006), the average WPM read during the winter and spring of first grade placed students at the 25th performance percentile and at elevated risk of future reading difficulty. However, with an average 6 WPM gain over the summer, students entered second grade reading 7 WPM more than the 25 WPM benchmark that identifies the fall of second-grade first performance quartile. Over the winter and spring of the second-grade academic year, absolute and relative performance levels continued to improve. The 78 WPM read on average in the spring of second grade placed the student sample just below the 10 WPM lower bound of the 50th percentile score (89 WPM) that reflects normative and expected performance (Hasbrouck & Tindal, 2006).

Analytic Procedures

The delivery of instruction at three different times across 2 academic years and the intervening summer along with the withdrawal of instruction in the two breaks before and after summer school define an RT design that is consistent in form with interrupted time-series (Bloom, 2003; Campbell, 1969; Shadish et al., 2002) and single-subject research (Horner et al., 2005; Kratochwill & Levin, 1992) designs. In RT contexts, fundamental changes in the outcome of interest are often expected to be discontinuous in reaction to the introduction or removal of some causally linked covariate (e.g., instruction). Conventional growth modeling techniques that estimate a single trajectory for the growth function over all time points are thus not appropriate (Gordon & Heinrich, 2004; Osgood & Smith, 1995; Singer & Willett, 2003). Instead, an alternative model specification is needed that represents the time-dependent, episodic treatment schedules and allows modeling of program, treatment, or intervention effects.

A modeling solution that captures the complexity of RT-based designs is offered by the procedural and analytic flexibility of piecewise growth models (PGM). The PGM is useful in the RT design context as individual changes in performance level and growth coordinated with the exact time at which intervention is delivered (or removed) can be estimated. By breaking the time function into pieces, separate parameters that explicitly represent distinct segments of the overall growth trajectory can be specified and tests for discontinuities between growth-trajectory components can be conducted (Singer & Willett, 2003). Of particular importance for the current study is the flexibility that PGM afforded for estimating and modeling the discontinuous changes in learning that were hypothesized to occur over the course of 2 academic years, in conjunction with the implementation of summer school, and between the breaks before and after the summer intervention period.

In order to estimate learning rates within and between different instructional and noninstructional periods and identify the impact of the summer instructional supplement, a two-level PGM with three segments in one case and with five segments in the second case was applied to the TORF time series. For both the three- and the five-segment models, two-level PGM (observations within students) were estimated without an additional third level because students changed classrooms each year and summer school was delivered to all students at one central site. All models were estimated with HLM, version 7 (Raudenbush et al., 2011) using restricted maximum-likelihood estimation.

Three-segment model. The three-segment model was composed of (*a*) a first-grade fall, winter, and spring segment (G1), (*b*) a segment capturing the change in performance across the entire summer (SUM, the change between the spring assessment in grade 1 and the fall assessment in grade 2), and (*c*) a second-grade fall, winter, and spring segment (G2). The three-segment model was used to estimate the change in summer learning in a manner consistent with much of the summer learning/summer school literature. The three-segment model also served as a baseline for evaluating the academic year and summer learning estimates obtained from the five-segment model.

The coding used to define the three-segment model's growth function is presented in Table 2 and is comparable to methods described in texts (Raudenbush & Bryk, 2002; Singer & Willett, 2003) and applications in the literature (Gordon & Heinrich, 2004; Hindman, Cromley, Skibbe, & Miller, 2011). Changes in oral reading fluency were estimated across the 92-week study period (i.e., 40-week first-grade academic year + 12-week summer "vacation" period + 40-week second-grade academic year = 92 weeks). The number of elapsed weeks associated with each of six measurement points during the 2 academic years was identified within this time span (grade 1: weeks 0, 18, 36; grade 2: weeks 52, 70, and 88). As shown in Table 2, two alternative parameterizations of the time function (i.e., differential rate and incremental coding) were used to obtain specific model estimates and statistical tests of interest.

On the far left side of Table 2, the three-segment differential rate coding scheme is presented. Differential rate coding was used to obtain actual parameter values and test whether the obtained estimates were statistically different from zero (e.g., Gordon & Heinrich, 2004; Raudenbush & Bryk, 2002). The differential rate coding scheme enabled direct estimation of the weekly gain in oral reading fluency across first grade (π_{1i}), the change in fluency associated with the 12-week summer period (π_{2i}), and the weekly gain in fluency across the second-grade academic year (π_{3i}) as specified in equation 1 below. Adjacent to the differential rate codes, the three-segment incremental coding scheme is presented. Increment codes were applied in order to specifically test whether the estimated change in reading fluency rate statistically differed from the baseline period (i.e., first grade). The incremental coding scheme produced an estimate of the weekly gain in oral reading fluency across first

Table 2. Differential Rate and Incremental Coding Scheme for Academic Year and Summer Changes in Reading Fluency for the Three- and Five-Segment Growth Models

	Three-Segment Model						Five-Segment Model									
	Differential Rate			Increment			Differential Rate					Increment				
Observation (Elapsed Weeks)	G1	SUM	G2	G1	SUM	G2	G1	SB1	SS	SB2	G2	G1	SB1	SS	SB2	G2
G1 mid-September (0)	−36	0	0	−36	0	0	−36	0	0	0	0	−36	0	0	0	0
G1 beg-January (18)	−18	0	0	−18	0	0	−18	0	0	0	0	−18	0	0	0	0
G1 mid-May (36)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SS mid-July (44)	–	–	–	–	–	–	0	1	0	0	0	8	1	0	0	0
SS end-July/beg-August (46)	–	–	–	–	–	–	0	1	2	0	0	10	1	2	0	0
SS mid-August (48)	–	–	–	–	–	–	0	1	4	0	0	12	1	4	0	0
G2 beg-September (52)	0	1	0	16	1	0	0	1	4	1	0	16	1	8	1	0
G2 beg-January (70)	0	1	18	34	1	18	0	1	4	1	18	34	1	26	1	18
G2 mid-May (88)	0	1	36	52	1	36	0	1	4	1	36	52	1	44	1	36

Note.—G1 = grade 1, SUM = summer, G2 = grade 2, SB1 = summer break 1, SS = summer school, SB2 = summer break 2.

grade (π_{1i}) and the change in fluency rate (relative to the first-grade baseline) associated with the summer period (π_{2i}) and the second-grade academic year (π_{3i}).

Under each of the parameterizations, time was represented both as a continuous variable indexing the number of elapsed weeks at each assessment point during each academic year (centered at the third measurement point, the grade 1 spring assessment) and as a categorical dummy variable (0, 1) capturing the difference in oral reading fluency from the final TORF assessment in first grade to the initial TORF assessment in second grade. With the coding used to represent the change in reading fluency within and between the different time periods, the status parameter (π_{0i}) was defined as the expected fluency rate in mid-May of first grade. Equations 1 and 2 specify the basic form of the two-level, three-segment PGM model whereby estimation of the growth function occurred at level 1 and modeling of interindividual variation in the fluency status and change outcomes occurred at level 2. Dummy-coded student and cohort membership characteristics (referent group for each contrast: male, White, economically advantaged, English proficient, general education, cohort 3) were entered as the level 2 predictors (a_{Pi}) of each of the status and change outcomes. In addition, number of days of summer school attendance was also included as a level 2 predictor of the summer change in reading fluency.

Level 1 (measurement occasions):

$$Y_{ti} = \pi_{0i} + \pi_{1i}(G1) + \pi_{2i}(SUM) + \pi_{3i}(G2) + e_{ti}. \tag{1}$$

Level 2 (students):

$$\begin{aligned} \pi_{0i} &= \beta_{p0} + \beta_{pi}(a_{Pi}) + r_{0i} \\ \pi_{pi} &= \beta_{p1} + \beta_{pi}(a_{Pi}) + r_{1i}, \text{ for each slope parameter.} \end{aligned} \tag{2}$$

With application of the two different coding schemes, it is important to note that two of the four parameters that were estimated at level 1 and modeled by student predictor variables at level 2 were conceptually distinct. In the differential rate model, academic year growth and summer change values were directly estimated and modeled, whereas in the incrementally coded model, estimated changes in summer and second-grade fluency rate relative to baseline were regressed on the set of predictor variables. In both models, the end-of-first-grade status (π_{0i}) and rate of first-grade fluency growth outcomes (π_{1i}) were equivalent.

Five-segment model. In the five-segment model, the course of learning across each of the in- and out-of-school periods was estimated, including both summer breaks and during summer school. The five segment "A-B-A-B-A" RT model was composed of: (*a*) a segment estimating growth during first grade (G1; fall, winter, and spring), (*b*) a segment representing the change in performance between the end of first grade and the beginning of summer school (SB1; the change between the spring assessment in grade 1 and the first summer school assessment), (*c*) a segment estimating growth during summer school (SS), (*d*) a segment representing the change in performance between the end of summer school and the start of second grade (SB2, the change between the final summer school assessment and the fall assessment in grade 2), and (*e*) a segment estimating growth during second grade (G2; fall, winter, and spring). The five-segment model specification was thus closely aligned with the complex RT design that characterized the annual pattern of instruction for students who attended summer school. The five-segment model was used to isolate the change in reading fluency between the breaks before and after summer school and to more accurately identify the amount of learning that occurred during the summer school intervention.

The coding associated with the five-segment model was more complex due to the need to specify additional segments to represent the RT design. The presentation of coding methods for more complex PGM models is less common. However, one of the advantages of PGM is the flexibility in accommodating varying times and intervals of measurement. As shown on the right side of Table 2, changes in oral reading fluency were again estimated across the 92-week study period, but with the addition of the three summer measurement occasions (weeks 44, 46, 48). Expansion of the differential rate and incremental coding schemes provided estimates of the actual gain in fluency during each period and the relative change in fluency rates between contiguous segments, respectively. Time was again represented as a continuous variable indexing the number of elapsed weeks at each assessment point during each schooling period and as categorical dummy variables capturing the change in oral reading fluency over the two summer breaks. By virtue of the coding scheme associated with each model, the status parameter (π_{0i}) was again defined as the fluency rate at the end of first grade.

Equations 3 and 4 specify the five-segment PGM. Using the same general structure underlying the three-segment model, within-person growth was modeled at level 1 and between-person variability in status and the absolute and relative change outcomes was modeled by the set of student predictor and cohort membership variables at level 2.

Level 1 (measurement occasions):

$$Y_{ti} = \pi_{0i} + \pi_{1i}(G1) + \pi_{2i}(SB1) + \pi_{3i}(SS) + \pi_{4i}(SB2) + \pi_{5i}(G2) + e_{ti}. \quad (3)$$

Level 2 (students):

$$\pi_{0i} = \beta_{p0} + \beta_{pi}(a_{Pi}) + r_{0i}$$
$$\pi_{pi} = \beta_{p1} + \beta_{pi}(a_{Pi}) + r_{1i}, \text{ for each slope parameter.} \quad (4)$$

In the five-segment model, the first "summer" break spans 8 weeks (mid-May to mid-July) and the second summer break spans 4 weeks (mid-August to mid-September), whereas the summer period in the three-segment model spanned 16 weeks. As a result, the three- and five-segment "summer" estimates are not directly comparable as the three-segment estimates represent either the total or relative change in reading fluency from mid-May of grade 1 to the beginning of grade 2, while the five-segment estimates represent either the weekly growth rate or the relative increment in reading fluency growth during the 5-week summer literacy intervention.

Results

Three-Segment Model

Table 3 presents the results of the three- and five-segment models. Estimation of the three-segment differential rate model revealed that on average student reading

Table 3. Three- and Five-Segment Model Results (N = 250)

	Three-Segment Model					Five-Segment Model			
	Differential Rate		Increment			Differential Rate		Increment	
	Mean (SE)	Variance	Mean (SE)	Variance		Mean (SE)	Variance	Mean (SE)	Variance
Intercept, π_{0i}	25.61 (.77)*	119.82	25.61 (.77)*	119.86	Intercept, π_{0i}	25.66 (.78)*	129.67	25.66 (.78)*	129.67
G1, π_{1i}	.64 (.03)*	.10	.64 (.03)*	.10	G1, π_{1i}	.65 (.03)*	.11	.65 (.03)*	.11
					SB1, π_{2i}	.42 (.36)	3.10	−4.76 (.47)*	6.98
SUM, π_{2i}	5.92 (.62)*	33.06	−4.34 (.68)*	4.30	SS, π_{3i}	1.92 (.12)*	1.23	1.27 (.11)*	.69
					SB2, π_{4i}	−2.35 (.42)*	1.21	−10.03 (.71)*	13.91
G2, π_{3i}	1.24 (.03)*	.17	.60 (.03)*	.20	G2, π_{5i}	1.24 (.03)*	.18	−.68 (.11)*	1.08

Note.—G1 = grade 1, SUM = summer, G2 = grade 2, SB1 = summer break 1, SS = summer school, SB2 = summer break 2; SE = standard error; three-segment model deviance = 9,033.35, parameters estimated = 11, five-segment model deviance = 13,350.27, parameters estimated = 22.

* $p < .05$.

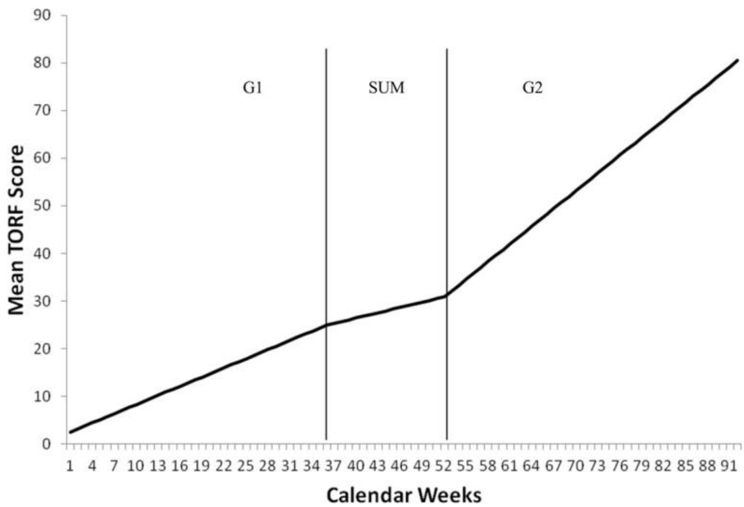

Figure 1. Estimated three-segment TORF growth trajectory.

fluency increased almost two-thirds of a word per week in first grade ($\pi_{1i} = 0.64$). The academic year gain translated into an average of approximately 26 WPM fluency rate toward the end of first grade ($\pi_{0i} = 25.61$). During the "summer" (which included the final 4 weeks of first grade and 5 weeks of summer school), fluency growth remained positive as students were able to read 6 more WPM ($\pi_{2i} = 5.92$) at the start of second grade than toward the end of first grade, a gain of .37 words per week over the 16-week period bounding the first-grade mid-May assessment and the mid-September assessment at the start of second grade (i.e., 5.92/16 = .37). Fluency growth was greater during second grade, as students demonstrated gains of one and a quarter words per week ($\pi_{3i} = 1.24$), essentially doubling their first-grade fluency acquisition rate.

To evaluate whether the observed between-period differences in the academic year and summer oral reading fluency rates were statistically significant, the three-segment model was reestimated using increment codes. The incremental coding enabled direct tests of whether the rate of fluency gains in the summer and second grade differed from that observed during the first-grade baseline period. The weekly gain in reading fluency associated with the summer period was statistically lower than the rate of gain observed during first grade (i.e., 0.37 vs. 0.64 WPM/week, $\pi_{2i} = -4.34$, $p < .05$; $-4.34/16 = -0.27$, $0.64 - .0.27 = 0.37$). The test of whether there was an increment in the growth rate from grade 1 to grade 2 showed that second-grade growth was statistically greater than the growth in first grade (i.e., 1.24 vs. 0.64 WPM/week, $\pi_{3i} = 0.60$, $p < .05$; $0.60 + 0.64 = 1.24$). Multiparameter variance component tests revealed that differential rate and incrementally coded models that allowed each growth trajectory component to vary randomly (deviance = 9,033.35, parameters estimated = 11) provided a better fit to the data than models that constrained growth components to be equivalent across students (deviance = 9,862.99, parameters estimated = 2), $\chi^2 \Delta (9) = 829.65$, $p < .05$. Figure 1 presents the trajectory of oral reading fluency growth across the study period. In the figure, it can be seen that growth was positive in each segment, including over the time coincident with the summer vacation period.

When covariates were added to the differential rate model, gender, ethnicity, students' poverty level, and English language status were not statistically significant predictors of any of the fluency status and change outcomes. However, relative to their general education peers, special education students grew at a rate .20 WPM/week slower during first grade and read 9 fewer WPM on average toward the end of the first-grade academic year. Special education students also had only a .50 WPM total gain in reading fluency over the summer and continued to grow at a rate .30 WPM/week slower than their peers during second grade. With respect to students in cohort 3, students in cohort 1 read approximately 5 fewer WPM at the end of first grade, but had an 8 WPM gain over the summer, while weekly grade 1 gains were .22 WPM/week larger for students in cohort 2. During the 16-week summer period, students with higher summer school attendance gained at a faster rate than students who attended less frequently. On average, for each additional day of attendance, students gained an additional .03 WPM in reading fluency/week. When covariates were added to the incrementally coded model, the pattern of results was analogous to the differential rate model with respect to relationships between student background and cohort membership and the first-grade status and growth outcomes. However, the change in fluency during the summer (relative to grade 1 and cohort 3 performance) was more modest for students in cohort 1 and cohort 2, while in grade 2 the learning increment was lower for cohort 2 students. Translated into a weekly gain metric, while cohort 3 students were gaining .27 WPM/week less than in grade 1, the summer rate of gain was only .09 and .20 WPM/week less for cohort 1 and cohort 2 students, respectively. In grade 2, while cohort 3 students were gaining 1.42 WPM/week, conditional weekly gains were 1.36 WPM for cohort 1 and 1.27 WPM for cohort 2.[2]

Five-Segment Model

The more complex course of change over time associated with the five-segment model (three periods of instructional engagement: first grade, summer school, and second grade; and two out-of-school periods: the summer breaks before and after summer school) was examined next. Using the differential coding scheme presented in Table 2, the academic year estimates were within a decimal place of their counterparts in the three-segment model: end of first-grade status ($\pi_{0i} = 25.66$); first-grade growth ($\pi_{1i} = 0.65$); second-grade growth ($\pi_{3i} = 1.24$). In addition, the estimates representing the summer period revealed that students gained approximately two words/week ($\pi_{2i} = 1.92$) while participating in the instructional intervention but stagnated during the first summer break ($\pi_{4i} = 0.42$) and lost ground during the second summer break ($\pi_{5i} = -2.35$). The positive rate of fluency growth during summer school and the negative change in reading-fluency status during the second summer break were statistically different from zero, whereas the estimated change in fluency during the first summer break was not.

To evaluate whether changes between periods were statistically significant, the five-segment model was reparameterized using increment codes. Estimation of the increment model revealed that the rate of growth associated with the summer school period was statistically greater than the rate of growth during first grade (1.92 vs. 0.65 WPM/week, $\pi_{2i} = 1.27, p < .05; 1.27 + 0.65 = 1.92$) as well as the rate of growth during second grade (1.92 vs. 1.24 WPM/week, $\pi_{3i} = -0.68$, $p < .05; 1.92 - 0.68 = 1.24$). With respect to the two out-of-school summer break periods, the weekly gain in reading fluency associated with the first summer break was statistically lower than

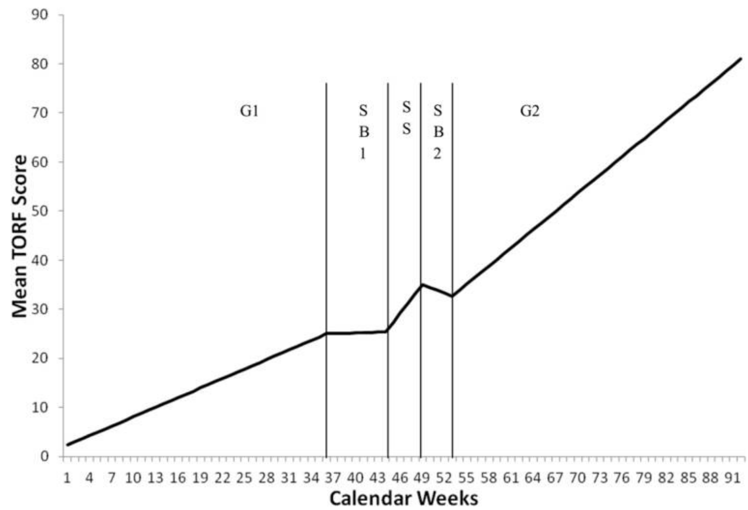

Figure 2. Estimated five-segment TORF growth trajectory.

the rate of gain observed during first grade (i.e., 0.05 vs. 0.65 WPM/week, $\pi_{4i} = -4.76$, $p < .05$; $-4.76/8 = -0.60$, $0.65 - .0.60 = 0.05$) and the weekly decline in fluency during the second summer break directly contrasted with the 2 WPM/week gain observed in conjunction with summer school (i.e., −0.59 vs. 1.92 WPM/week, $\pi_{5i} = -10.03$, $p < .05$; $-10.03/4 = -2.51$, $1.92 - 2.51 = -0.59$). Multiparameter variance component tests revealed that differential rate and incrementally coded models that allowed each growth trajectory component to vary randomly (deviance = 13,350.27, parameters estimated = 22) provided a better fit to the data than models that constrained growth components to be equivalent across students (deviance = 14,720.23, parameters estimated = 2), $\chi^2\,\Delta\,(20) = 1{,}369.96$, $p < .05$. Figure 2 depicts the course of learning for summer school attendees across the five time periods. In the figure, it can be seen that while the rate of reading fluency was increasing during each schooling period, the trend in performance was either flat or negative during the out-of-school segments before and after summer school.

When covariates were added to the five-segment differential rate model, gender, ethnicity, students' poverty level, and English language status were not statistically significant predictors of any of the fluency status and change outcomes. However, special education students were again observed to grow at a rate .20 WPM/week slower than their peers during first grade and read approximately 9 fewer WPM on average at the end of the first grade. Special education students also gained .76 WPM/week and .28 WPM/week less than their peers during the summer intervention period and over second grade, respectively. Cohort performance during first and second grade was analogous to that reported in the three-segment model, as was the positive relationship between summer school attendance and fluency growth. On average, for each day of attendance, students produced an additional .13 WPM per week of summer school. No other relationships between predictors and the summer break outcomes were observed, with the exception of the 3 WPM/week less negative second summer break decline in fluency for cohort 2 students. With respect to the incrementally coded model, first-grade status and growth results were analogous to the differential model. During the first summer break, cohort 1 had less and cohort 2 had more of a decrement in fluency (relative to grade 1 and cohort 3), while during the second break only the lower decrement in fluency for cohort 1 students (relative to summer school and cohort 3) was statistically significant. Translated into a weekly gain metric, while cohort 3 students were conditionally gaining .76 WPM/week less than in grade 1 (i.e., a .10 WPM/week loss), the rate of gain during the first summer break was .48 WPM/week less (i.e., a .20 WPM/week gain) and 1.07 WPM/week more (i.e., a .40 WPM/week loss) for cohort 1 and cohort 2 students, respectively. During the second summer break, cohort 1 students' learning decrement (relative to summer school and cohort 3) was 2.01 WPM/week (i.e., a .53 WPM/week loss during the 4-week break following summer school). None of the between-group or cohort differences in the increment/decrement to learning in any of the other schooling and nonschooling periods were statistically significant.

Supplemental Analysis

The piecewise estimates from the five-segment model demonstrated that positive gains in reading fluency occurred only during periods of schooling. However, it is important to reiterate that the first-grade spring assessment was not administered on the last day of the academic year. With the final academic year measurement point occurring in mid-May, the first summer break segment included 4 weeks of academic year instruction (i.e., mid-May to mid-June). As a result, the impact of the first "summer break" was expected to be underestimated due to the learning that occurred between the spring assessment period and the end of the academic year.

In an attempt to correct for this potential bias and estimate the "true" change in learning associated with the first summer break, 4 weeks of estimated first-grade reading fluency growth were added to the mid-May TORF scores to derive an end-of-year (mid-June) projected score for each student (i.e., $\pi_{0i} + [4 \times \pi_{1i}]$). Estimation of a five-segment model that included projected end-of-year scores revealed that the "true" change in fluency across the first 4-week summer break (i.e., mid-June to mid-July) was a -1.64 WPM loss ($SE = 0.21$, $p < .05$). Isolation of the change in reading fluency that occurred during the first summer break enables one to estimate how the summer school growth rate may have been affected had a last day/first day academic year measurement schedule been utilized. Separation of the two out-of-school summer break estimates from the summer school growth estimate (i.e., $4 \times 1.92 = 7.68$, $7.68 - 1.64 - 2.35 = 3.69$, $3.69/7.68 = .48$) revealed that the summer school growth rate would have been underestimated by approximately 50% had a last day/first day academic year-to-academic year measurement approach been used.[3] Figure 3 displays the average learning trajectory from the supplemental analysis, reflecting the zigzag pattern of development that arose in conjunction with the schooling and nonschooling periods.

Discussion

In the past decade, increasingly creative methods have been utilized to estimate in- and out-of-school learning rates and strengthen inferences regarding schooling effects (e.g., Downey et al., 2004, 2008; von Hippel, 2009). The trajectory of achievement growth established in the analysis of local and nationally representative data sets suggests that learning is generally positive during periods of instruction but slows or declines when students are not in school during the summer (Alexander et

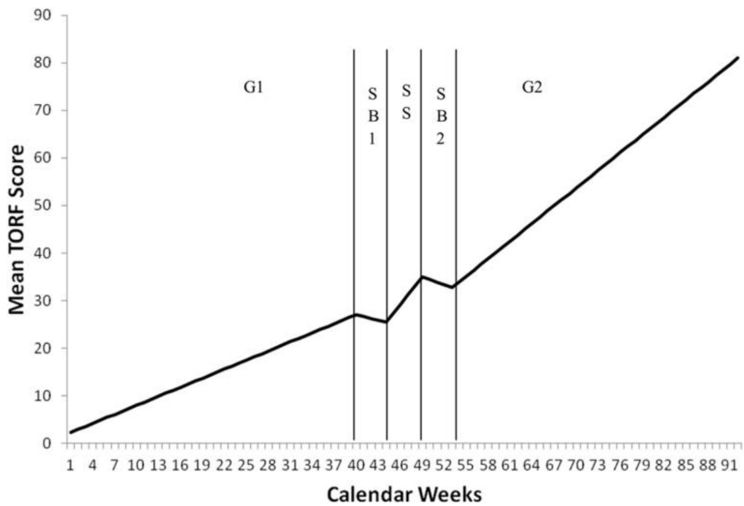

Figure 3. Estimated five-segment TORF growth trajectory using the projected end-of-first-grade score.

al., 2001; Bryk & Raudenbush, 1988; Downey et al., 2008; Raudenbush, 2004). However, the actual magnitude of the summer learning "slide" can be difficult to pinpoint. If summer learning estimates are based on assessment scores not obtained from a last day/first day academic year measurement schedule, the learning that accrues after the final assessment before summer and the first assessment following summer can be inadvertently included in a summer change estimate (Burkam et al., 2004; Cooper et al., 1996; Heyns, 1987; McCombs et al., 2011). A similar problem occurs in studies specifically designed to estimate the learning gain specifically attributable to summer school (Cooper et al., 2000). Whether based on a randomized or regression discontinuity design or a weaker quasi-experimental alternative, the use of academic year test scores to estimate school-based summer learning is undermined by an inability to accurately capture achievement gains as a result of measurement occasions that are not aligned with the onset and conclusion of summer instruction.

Inaccuracies in the estimation of summer learning and the practical difficulties associated with the implementation of strong experimental research designs suggest that a methodological and assessment approach that closely aligns with the episodic presentation and removal of formal schooling would further strengthen the summer school/summer learning literature. In the present study, schooling effects were evaluated using an RT design and PGM techniques applied to reading-fluency data obtained over 2 academic years and the intervening summer. Results demonstrated the accrual of gains in reading fluency during periods of schooling coupled with periods of stagnation and/or loss when students were not in school. The observed pattern of student performance suggests that for the struggling readers studied herein, instruction had a positive impact regardless of when in the calendar year it was delivered. Moreover, in comparison with the rate of fluency acquisition obtained during either academic year, fluency skills were gained at a statistically greater rate during summer school when lengthy and exclusive literacy instruction was delivered. To the extent that these results can be generalized to other academic domains and schooling contexts, it reasonably follows that efforts to eliminate summer school and/or reduce the number of school days or contact hours in response to budgetary pressures will negatively impact student learning and make the timely attainment of federal or local accountability goals more difficult.

The current application of an RT/PGM framework sought to demonstrate the trajectory of reading fluency growth for a sample of struggling readers who participated in a 5-week summer literacy program. In the three-segment model, results showed that students' reading fluency increased over each time period. However, without clear separation of the summer vacation break from summer school, results associated with the three-segment model are somewhat difficult to interpret as the observed changes in summer reading fluency could be attributed in part to instruction received during first grade and/or the background, motivational, or maturational characteristics of the student sample rather than the positive impact of summer instruction. In contrast, observed growth during the academic year and summer schooling periods coupled with the flattening or decline in fluency acquisition during the two summer breaks in the five-segment model more clearly demonstrated how the provision and removal of instruction affected student learning rates. In particular, it is the application of the RT design with repeated and directionally consistent contrasts in fluency outcomes between each in- and out-of-school period that serves to effectively rule out several common threats to internal validity (e.g., history, maturation) and lend strength to the inference that instruction was associated with the observed academic year and summer fluency gains.

At the same time, the effect of capturing a portion of academic year learning in a summer break estimate was also apparent in the five-segment model results. With 4 weeks of academic year learning contained in the first summer break estimate, students were estimated to maintain their level of reading fluency during the period of time prior to the onset of summer school. This outcome differed with respect to the fluency loss observed during the 4-week break following the conclusion of summer school. However, after removing the gains in reading fluency that were likely attributable to the final 4 weeks of academic year instruction, the adjusted mid-June to mid-July (−1.64) summer break estimate was shown to be comparable in direction and size to that associated with the second 4-week summer break (−2.35). The impact of capturing a portion of the summer break in a summer school growth estimate was also used to highlight how identification of learning outcomes associated with a summer school program may be negatively impacted by a misaligned intervention and assessment schedule. In the present study, school-based summer reading fluency growth would have been underestimated by 50% (about a 1 vs. 2 WPM/week gain) had estimates been based on assessment scores obtained on the last day/first day of adjacent academic years instead of the time period in which summer instruction was actually implemented, a decidedly more modest estimate of program impact.

Consideration of study limitations is necessary to contextualize these results. First, as with many other summer school evaluations, students were offered supplementary instruction on a need basis, and participation was voluntary. Given that results were obtained from an administrative- and self-selected sample of early struggling readers located in a moderate-sized Pacific Northwest school district, broad conclusions should not be drawn regarding in- and out-of-school learning patterns and the potential strength of summer instruction delivered to students from different backgrounds, grades, and/or other scholastic contexts. Second, the absence of data on the educational resources available to students over the summer between first and second grade (see Burkam et al., 2004; Kim & White, 2008) and on the fidelity with which literacy instruction was delivered during each schooling period prevented a more nuanced examination of the trajectory of academic year and summer reading growth. Third, with the availability of only three observations during each instructional period, the change in reading fluency was necessarily modeled as a function of a linear growth segment. To the extent that performance deviated from an incremental trajectory, the addition of 4 weeks of linear first-grade growth (i.e., $\pi_{0i} + [4 \times \pi_{1i}]$) to derive an end-of-year projected score for the supplemental analyses may have resulted in either an over- (if decelerating; e.g., Nese et al., 2013) or understatement (if accelerating; e.g., Kim et al., 2010) of the decline in performance during the first summer break. Fourth, despite the RT within-group design and the patterned reversal in fluency outcomes, the lack of a control group prevented the establishment of a between-group comparison. Had a control group been included in the design, period-specific performance trends could be compared and inferences regarding the impact of summer school would be further strengthened.

Another set of limitations surround the study's reading fluency measure. In particular, it should be recognized that examination of performance on outcomes (e.g., phonemic awareness, vocabulary acquisition) more or less sensitive to practice effects and the instructional treatment may have resulted in different summer learning estimates. It should also be noted that despite substantial empirical support for strong relationships between reading fluency and reading comprehension measures (e.g., Baker et al., 2008; Crawford et al., 2001; Good et al., 2001; Nese et al., 2011; Reschly et al., 2009), results reported here may present an incomplete picture of literacy growth and development defined more broadly. Moreover, as the TORF forms used to measure reading fluency were designed to be of equivalent difficulty within and between first and second grade (Children's Educational Services, 1987), but were not explicitly equated or vertically linked, it is also possible that estimates reported here are flawed by form effects (Betts, Pickart, & Heistad, 2009; Francis et al., 2008). Form-to-form differences in passage difficulty can serve to undermine the accuracy of point estimates and the stability of growth estimates (Francis et al., 2008). However, in the methods used for TORF administration, form effects were likely ameliorated to some extent by use of three forms on each measurement occasion and an operational TORF score based on the median number of words correct per minute across the three forms. Nonetheless, it is important to acknowledge that the reading fluency results presented depend on the psychometric characteristics of the particular assessment methods used.

Although these limitations suggest a need for additional research using different samples, measures, and design controls, the average weekly rate of gain observed during the first- and second-grade academic years was generally comparable to that reported in other recent investigations and summaries of the oral reading fluency literature (e.g., Kim et al., 2010; Tindal, Nese, Stevens, & Alonzo, 2013). Yet the absolute increase in reading fluency over the summer differed from the expected decline in summer performance specified in published ORF norms (Hasbrouck & Tindal, 2006) and with respect to the drop in fluency outcomes reported in studies that have examined the fluency outcomes of individual students over multiple academic years (e.g., Baker et al., 2008; Kim et al., 2010). The summer fluency gains instead align well with meta-analytic findings regarding the benefits of supplemental summer instruction (Cooper et al., 2000; Kim & Quinn, 2013) and other recent investigations that have used relatively rigorous designs to examine summer school effects (Borman et al., 2009; Borman & Dowling, 2006; Schacter & Jo, 2005).

Growth effect sizes (Bloom, Hill, Black, & Lipsey, 2008) serve as an additional empirical benchmark for contextualizing our results. Effect sizes (ES) were calculated for each in- and out-of-school period by examining the mean difference from the beginning to the end of each period in ratio to the pooled standard deviation for the two time points (e.g., G1: fall to spring grade 1, SB1: spring grade 1 to the beginning of summer school, SS: beginning to end of summer school, etc.). To provide comparability for the growth effect sizes that are based on periods of differing duration, each ES is also expressed as mean change per week (in parentheses). Effect sizes during each in- and out-of-school period were G1: 1.68 WPM (0.047/week); SB1: 0.01 WPM (0.001/week); SS: 0.58 WPM (0.146/week); SB2: −0.146 WPM (−0.036/week); and G2: 2.06 WPM (0.057/week). Effect size estimates indicate that the summer instructional intervention was able to deliver standardized learning gains that exceeded those obtained during each academic year. The ES estimates also sharply highlight the contrast in learning between in- and-out of school periods. However, it should be emphasized that the within-group, growth effect size estimates do not directly speak to the relative (between-group) efficacy of the summer literacy intervention as an equivalent group of nontreated struggling readers was not included in the design.

In many contexts researchers wish to study the effects of programs, interventions, or other time-linked phenomena that occur periodically. The episodic introduction and removal of treatment may occur by design in controlled research settings or in conjunction with a naturally occurring event like the enactment/repeal of a law or policy. In the current study, the structure of the K–12 academic calendar, where 9 months of schooling alternates with 3 months of summer vacation, was leveraged to investigate summer school effects and in- and out-of-school learning patterns. Although specifically focused on the investigation of academic year and summer learning, the RT/PGM framework in which the study was situated can readily be applied to a wide variety of problems in the analysis of change and program evaluation, including response-to-intervention and single-subject research contexts where presentation, withdrawal, or change in treatments may occur repeatedly over the course of time. In these and similar field-based settings where it is not practical or ethical to withhold treatment from those in need, but where the type and timing of treatment or programmatic changes are recorded and key outcomes are contiguously measured, the piecewise modeling of RT design data is a useful strategy for increasing the validity of inferences regarding the impact of an intervention as well as in ascertaining the degree of individual responsiveness to a course of supplemental treatment.

Notes

Keith Zvoch is associate professor and Joseph J. Stevens is professor in the Department of Educational Methodology, Policy, and Leadership at the College of Education of the University of Oregon in Eugene. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A090369 awarded to the University of Oregon. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Address all correspondence to Keith Zvoch; e-mail: kzvoch@uoregon.edu.

1. Reading fluency measures are generally not administered until the winter of first grade, although different states and districts sometimes use assessment schedules that include a fall of first grade assessment (e.g., Kim et al., 2010). In the district where the current study is based, literacy teams at individual schools have flexibility to assess students' reading fluency at the beginning of first grade if performance on other early literacy indicators is deemed adequate or additional diagnostic data are needed. Of the 250 students in the study sample, 47 (19%) completed a TORF assessment upon entry to first grade.

2. For brevity, with the large number of statistically nonsignificant findings, coefficients for student predictor variables are not presented. Conditional model results are available from the authors upon request.

3. The projected mid-June score for each student (i.e., $\pi_{0i} + [4 \times \pi_{1i}]$) was also entered in the three-segment model. Estimation of a three-segment model that included projected end-of-year scores revealed that the "true" change in fluency across the 12-week summer period (i.e., mid-June to mid-September) was a 3.82 WPM gain ($SE = 0.47, p < .05$), a reduction of 35% from the original model estimate that included the final 4 weeks of first grade (i.e., 3.82/5.92 = .65). The estimated 3.82 WPM fluency gain derived from the three-segment model closely matched the 3.69 WPM fluency gain obtained by separating the two summer break estimates from the five-segment model's (with projected mid-June scores) summer school growth rate (i.e., $4 \times 1.92 = 7.68, 7.68 - 1.64 - 2.35 = 3.69$).

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, **23**(2), 171–191.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, **72**, 167–180.

America After 3PM. (2010). *Special report on summer: Missed opportunities, unmet demand*. New York: Wallace Foundation.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, **37**(1), 18–37.

Benson, J., & Borman, G. D. (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.

Betts, J., Pickart, M., & Heistad, D. (2009). An investigation of the psychometric evidence of CBM-R passage equivalence: Utility of reliability statistics and equating alternate forms. *Journal of School Psychology*, **47**, 1–18.

Bloom, H. S. (2003). Using "short" interrupted time series analysis to measure the impacts of whole-school reforms: With applications to a study of accelerated schools. *Evaluation Review*, **27**, 3–49.

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**, 289–328.

Borman, G. D., & Dowling, N. M. (2006). The longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial. *Educational Evaluation and Policy Analysis*, **28**, 25–48.

Borman, G. D., Goetz, M. E., & Dowling, N. M. (2009). Halting the summer achievement slide: A randomized field trial of the KindergARTen summer camp. *Journal of Education for Students Placed at Risk*, **14**, 133–147.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical model. *American Journal of Education*, **97**, 65–108.

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**, 1–31.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24, 409–429.

Children's Educational Services. (1987). *Test of Oral Reading Fluency (TORF)*. Eden Prairie, MN: Author.

Cooper, H., Batts Allen, A., Patall, E. A., & Dent, A. L. (2010). Effects of full-day kindergarten on academic achievement and social development. *Review of Educational Research*, **80**(1), 34–70.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, **65**(1), Serial No. 260.

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), 227–268.

Cooper, H., Valentine, J. C., Charlton, K., & Melson, A. (2003). The effects of modified school calendars on student achievement and on school and community attitudes. *Review of Educational Research*, **73**(1), 1–52.

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*, **7**, 303–323.

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.

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

Fairchild, R., Smink, J., & Stewart, A. B. (2009). *It's time for summer: An analysis of recent policy and funding opportunities*. Baltimore, MD: National Summer Learning Association.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*, **46**, 315–342.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, **5**, 239–256.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades; A practice guide.* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61–88). New York: Guilford.

Good, R. H., Kaminski, R. A., Shinn, M. R., Bratten, J., Shinn, M., Laimon, D., . . . Flindt, N. (2004). *Technical adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development* (Technical Report No. 7). Eugene: University of Oregon.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, **5**, 257–288.

Gordon, R. A., & Heinrich, C. J. (2004). Modeling trajectories in social program outcomes for performance accountability. *American Journal of Evaluation*, **25**(2), 161–189.

Hasbrouck, J., & Tindal, G. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *Reading Teacher*, **59**, 636–644.

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

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development*, **58**, 1,151–1,160.

Hindman, A. H., Cromley, J. G., Skibbe, L. E., & Miller, A. L. (2011). Conventional and piecewise growth modeling techniques: Applications and implications for investigating Head Start children's early literacy learning. *Evaluation Review*, **35**(3), 204–239.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education, *Exceptional Children*, **71**(2), 165–179.

Individuals with Disabilities Education Improvement Act, U.S.C. H.R. 1350 (2004).

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, **86**(1), 226–244.

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*, **83***(3)*, 386–431.

Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading*, **12**(1), 1–23.

Kim, Y. S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology*, **102**(3), 652–667.

Kratochwill, T., & Levin, J. R. (1992). *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Erlbaum.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, **76**(2), 275–313.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York: Guilford.

Matsudaira, J. D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics*, **142**, 829–850.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, **98**(1), 14–28.

McCombs, J. S., Augustine, C. H., Schwartz, H. L., Bodilly, S. J., McInnis, B., Lichter, D. S., & Cross, A. B. (2011). *Making summer count: How summer programs can boost children's learning*. Santa Monica, CA: RAND.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum based measurement to predict performance on state assessments in reading. *School Psychology Review*, **33**, 193–203.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Rockville, MD: National Institutes of Health (NIH Pub. No. 00-4754).

Nese, J. F. T., Biancarosa, G., Cummings, K., Kennedy, P., Alonzo, J., & Tindal, G. (2013). In search of average growth: Describing within-year oral reading fluency growth across grades 1–8. *Journal of School Psychology*, **51**, 625–642.

Nese, J. F. T., Park, B. J., Alonzo, J., & Tindal, G. (2011). Applied curriculum-based measurement as a predictor of high-stakes assessment: Implications for researchers and teachers. *Elementary School Journal*, **111**, 608–624

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Osgood, D. W., & Smith, G. L. (1995). Applying hierarchical linear modeling to extended longitudinal evaluations: The Boys Town follow-up study. *Evaluation Review*, **19**(1), 3–38.

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

Raudenbush, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Paper presented at the William H. Angoff Memorial Lecture Series, Princeton, NJ. Retrieved from http://www.ets.org/research/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, A. S., Fai, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, **47**, 427–469.

Schacter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, **28**(2), 158–169.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools*, **43**, 527–536.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: Oxford University 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.

Slocum, T. A., Street, E. M., & Gilbert, G. (1995). A review of research and theory on the relation between oral reading rate and reading comprehension. *Journal of Behavioral Education*, **5**, 377–398.

Smink, J. (2011). A new vision for summer school. *Educational Leadership*, **69**(4), 64–67.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Rep. 109). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Nese, J. F. T., Stevens, J., & Alonzo, J. (2013). *Within-year achievement growth trajectories using progress monitoring measures*. Paper presented at the annual meeting of the National Council on Measurement in Education.

von Hippel, P. T. (2009). Achievement, learning, and seasonal impact as measures of school effectiveness: It's better to be valid than reliable. *School Effectiveness and School Improvement*, **20**(2), 187–213.

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment*, **11**, 85–104.

Zvoch, K., & Stevens, J. J. (2011). Summer school and summer learning: An examination of the short and longer term changes in student literacy. *Early Education and Development*, **22**(4), 649–675.

Zvoch, K., & Stevens, J. J. (2013). Summer school effects in a randomized field trial. *Early Childhood Research Quarterly*, **28**(1), 24–32.