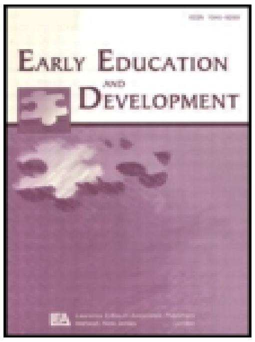

Early Education and Development

ISSN: 1040-9289 (Print) 1556-6935 (Online) Journal homepage: http://www.tandfonline.com/loi/heed20

Summer School and Summer Learning: An Examination of the Short- and Longer Term Changes in Student Literacy

Keith Zvoch & Joseph J. Stevens

To cite this article: Keith Zvoch & Joseph J. Stevens (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, DOI: 10.1080/10409289.2010.489891

To link to this article: http://dx.doi.org/10.1080/10409289.2010.489891

Published online: 08 Aug 2011.

EARLY EDUCATION AND DEVELOPMENT, *22*(4), 649–675

ISSN: 1040-9289 print/1556-6935 online
DOI: 10.1080/10409289.2010.489891

Summer School and Summer Learning: An Examination of the Short- and Longer Term Changes in Student Literacy

Keith Zvoch and Joseph J. Stevens
Department of Educational Methodology, Policy, and Leadership, University of Oregon

Research Findings: Multiple student cohorts were longitudinally tracked and student participation in a summer program bridging the 1st- and 2nd-grade academic years was recorded to examine selection and efficacy issues related to a summer school implementation in the Pacific Northwest. The estimation of regression discontinuity models uncovered evidence of a local average treatment effect. At the cutscore for program admission, participating students had estimated summer oral reading fluency gains approximately 0.40 *SD* larger than those of nonparticipants. Further examination of the literacy outcomes among the sample of cutscore eligible students revealed that struggling readers who participated in the summer program increased their level of reading fluency relative to struggling readers who declined an invitation to participate. However, the advantage gained by cutscore eligible participants was not sustained over the subsequent academic year. *Practice or Policy:* These results suggest that supplemental summer instruction delivered to at-risk students may promote literacy gains during the otherwise challenging summer months and thereby serve as a useful intermediary tool for K–12 stakeholders seeking to keep struggling readers on track toward proficiency.

Accountability pressures stemming from the No Child Left Behind (NCLB; No Child Left Behind Act, 2002) federal legislation have promoted the use of initiatives that restructure or extend the academic day (e.g., smaller learning

Correspondence regarding this article should be addressed to Keith Zvoch, Department of Educational Methodology, Policy, and Leadership, College of Education, University of Oregon, 5267 University of Oregon, Eugene, OR 97403-5267. E-mail: kzvoch@uoregon.edu

communities, full-day kindergarten) or otherwise optimize the efficacy of daily instructional practice (e.g., Cotton, 2001; Lee, Burkam, Ready, Honigman, & Meisels, 2006; Lee & Smith, 2001; Walston & West, 2004). For many students, access to a longer and/or more effective instructional day aids in offsetting home- and community-based resource challenges. However, a limiting factor in the extent to which schools and school-based initiatives can directly promote student achievement success is the time that students spend outside of the scholastic environment (Downey, von Hippel, & Broh, 2004; Downey, von Hippel, & Hughes, 2008). Particularly salient for schools on traditional academic calendars is the 3-month respite from schooling that students experience each summer. The long summer break provides an additional burden for schools struggling to meet NCLB annual yearly progress objectives as summer learning slows (relative to school-year learning) for students of advantage and flattens or atrophies for disadvantaged or at-risk students (Alexander, Entwisle, & Olson, 2001; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). Awareness of the potential for the "summer slide" to negatively impact a school's ability to meet NCLB accountability standards has led educational leaders to design and implement programs to keep students progressing toward proficiency over the summer vacation period (Jacob & Lefgren, 2004; Matsudaira, 2008; von Hippel, 2007). However, the extent to which students voluntarily participate and the degree to which students and schools may benefit from the provision of a supplementary summer program is not completely known. An evaluation of one district's summer school initiative is presented here as a means of addressing some of the design and analytic issues that have been identified in previous topical studies and reviews (Austin, Roger, & Walbesser, 1972; Borman, 2000; Borman, Benson, & Overman, 2005; Borman & Dowling, 2006; Cooper, Charlton, Valentine, & Muhlenbruck, 2000). Particular attention is given to documenting the characteristics of students who accepted or declined a summer placement as well as accounting for potential selection effects in models designed to estimate the short- and longer term efficacy of the intervention.

The NCLB requirement that all students in all schools be content area proficient suggests that schools have responsibility for the learning that occurs during the school year as well as the learning (or lack of learning) that takes place over the summer. Responsibility for the learning that occurs outside of the traditional 9-month academic year has led school districts to consider alternative schedules (e.g., year-round calendars) or implement summer programs as a means of bridging the 3-month vacation period (von Hippel, 2007). The need to provide a summer scaffold stems from a body of literature and the common experience of educational professionals that students generally do not maintain a strong upward learning trajectory over the summer months. Educators readily observe that learning slows for students of relative advantage and often declines for students from disadvantaged backgrounds. The summer achievement decrement observed among disadvantaged students is far from inconsequential. Extensive longitudinal study of the achievement status and growth of a sample of students from the Baltimore Public Schools revealed that the ninth-grade achievement gap between advantaged and disadvantaged students could be directly tied to the cumulative effect of annual summer learning rate differences (Alexander et al., 2001). Moreover, the accrued differences were closely related to whether a student received a college preparatory placement, completed high school, or attended a 4-year college (Alexander, Entwisle, & Olson, 2007). The finding of a differential summer slide and negative educational outcomes in the Baltimore Public Schools is not unique. A similar pattern of results and consequences has been observed by other investigators who have conducted longitudinal examinations of the summer achievement outcomes of students using data from national-, state-, or district-level samples. On average, students lose approximately 1 month of grade-equivalent skills over the summer. When expressed as an effect size, the average difference between spring and fall scores is about a tenth of a standard deviation (Cooper et al., 1996). Compounded over time, qualitatively distinct seasonal learning rates present a recurring challenge to K–12 stakeholders. The dip or decline in learning trajectories requires that teachers spend valuable instructional time reviewing or reteaching prior content at the beginning of each school year. For schools, time spent covering material from the previous year is time not spent preparing students to meet the proficiency standards of the current year. The summer setback is thus an unwelcome barrier for students trying to increase their content area knowledge and schools struggling to meet NCLB annual yearly progress objectives.

One approach that has been used by school leaders to mitigate the summer setback is the adoption and use of a year-round academic calendar. A recent estimate indicated that 2,353 public elementary schools in the United States, or 3.5%, followed a year-round calendar in 2005 (von Hippel, 2007). Year-round schools provide the same number of instructional days as their traditionally scheduled 9-month counterparts but do so over the entire calendar year. The redistribution of the traditional 3-month summer vacation break across the 12-month calendar year yields more frequent but shorter instructional disruptions. The shorter and more balanced instructional breaks are seen as a means of better promoting the continued scholastic engagement that is requisite to offset summer achievement declines. However, evidence regarding the relative efficacy of year-round schedules has been mixed. Although children who attend school over the summer learn more than their peers who experience a traditional summer vacation,

Early Education and Development 2011.22:649-675.

year-round attendees learn less during the regular academic year (as more vacation days are experienced). Overall, the differential rates of growth balance, and students learn a similar amount over the entire year (Cooper, Valentine, Charlton, & Melson, 2003; McMillen, 2001; von Hippel, 2007). The evidence regarding year-round scheduling effects on economically disadvantaged or otherwise low-achieving students is slightly more positive. At-risk students in year-round schools tend to outperform their traditionally schooled peers by approximately 0.05 to 0.10 *SD* on common measures of academic achievement (Cooper et al., 2003; McMillen, 2001). Nonetheless, the modest gains in achievement that have been associated with the adoption of a year-round schedule may not always be sufficient to outweigh the attendant disruption that follows a switch to a 12-month schedule (Cooper, 2004; von Hippel, 2007).

An alternative to a formal lengthening of the academic schedule is to offer supplementary instruction through a targeted summer program. Instruction delivered to low-achieving youth and/or youth from disadvantaged backgrounds may serve to limit the summer loss or facilitate a summer gain by providing these children with educational resources that may not be generally available in the home and community (Borman et al., 2005; Borman & Dowling, 2006; Cooper et al., 2000; Schacter & Jo, 2005). The impetus for a targeted summer initiative generally follows from the "faucet theory" of learning. The faucet theory suggests that distinct seasonal learning patterns stem from the flow of educational resources available at different points in the calendar year (Alexandar et al., 2001; Entwisle, Alexander, & Olson, 1997). During the academic year, the educational resource faucet is turned on, allowing all children the opportunity to learn. However, the resource flow differs for students from divergent economic backgrounds in the summer when school is not in session. Whereas students of advantage continue to have access to a flow of home- and community-based educational resources, impoverished students experience a resource tap that is often diminished to a trickle. Differential access to books and educational materials, educational enrichment programs and opportunities, and high educational expectations are thought to create the learning inequities that routinely manifest in the summer (Entwisle & Alexander, 1997; Heyns, 1978, 1987; Kim, 2004). To offset the unequal distribution and availability of educational resources, targeted summer school programs attempt to provide students with similar or enhanced access to the scholastic conditions that enable academic year success. Key to these initiatives is an academic focus and continued constructive engagement of students. More specifically, an organizational structure that provides a small learning community environment and direct instruction that makes effective use of all available instructional time is

Early Education and Development 2011.22:649-675.

requisite to keep the educational resource faucet flowing (Alexander et al., 2007; Borman & Dowling, 2006; Cooper et al., 2000).

A comprehensive review of summer school programs revealed the challenges associated with evaluating their efficacy but also indicated that supplementary instruction delivered over the summer may provide an academic benefit to children (Cooper et al., 2000). Summer school attendance generally enables economically advantaged students to maintain an upward learning trajectory and prevents economically disadvantaged students from experiencing the commonly observed summer decline. In particular, when considered relative to that of students with similar demographic backgrounds, the performance of economically disadvantaged summer school students was estimated at approximately 0.20 *SD* higher across a range of achievement test outcomes (Cooper et al., 2000). For students from middle-class backgrounds, summer school effects were greater. Effect size estimates of the difference between middle socioeconomic status treatment and control students were approximately one half of a standard deviation. Overall, Cooper and colleagues (2000) concluded that "the positive effect of summer school was about one quarter of a standard deviation" (p. 90), an impact on student achievement similar to that observed in programs with similar goals implemented during the course of a regular school year. Yet these authors also cautioned that many of the summer school programs included in the review provided a relatively weak form of treatment to students. The provision of remedial instruction or instruction disconnected from the academic year curriculum may have led to an underestimation of the potential benefits of more academically intensive summer programs.

The literatures on summer learning rates and supplementary summer instruction shine light on the challenges and opportunities the summer vacation period presents to K–12 stakeholders. These literatures also highlight a number of thorny methodological issues that can undermine the validity of inferences drawn from the study of summer learning outcomes. Of particular importance are issues associated with treatment assignment and participation. Rather than being assigned to summer school on a randomized basis, students offered a summer placement tend to be specifically identified as more economically disadvantaged and/or less academically proficient than their peers. Eligible students may also be more or less inclined or able to participate in a summer program. Together, the manner in which students are selected to participate and whether selected students actually participate in a voluntary summer program create a host of selection-based validity threats that can undermine the integrity of the inferences drawn from a counterfactual test (see Shadish, Cook, & Campbell, 2002). The selection issues identified in the summer school literature suggest that research designs that (a) control or better account for the need-based selection and voluntary participation of students, (b) track the longitudinal achievement progress of students pre- and posttreatment, and (c) include examination of the performance of treatment and control students from multiple cohorts are needed to further elucidate the manner and degree to which supplemental summer instruction can serve to offset the summer slide experienced by at-risk students (Borman, 2000; Borman et al., 2005; Borman & Dowling, 2006; Cooper et al., 2000).

In light of the practical importance and methodological challenges presented by the investigation of school-based summer initiatives, the current study was designed to investigate selection and efficacy issues related to a summer school implementation in a moderately sized school district in the Pacific Northwest. To evaluate the short- and longer term impact of the program, multiple student cohorts were longitudinally tracked over portions of two academic years, and student eligibility and participation status were identified. A series of inferential models was applied to literacy assessment scores. Regression discontinuity (RD) models were used to estimate the change in summer literacy performance for attendees, invitees, and their counterparts at the cutscore used for offering a summer school placement. Piecewise growth models were then used to compare the literacy growth of cutscore eligible program attendees and their cutscore eligible peers who declined to participate. Possible selection effects in the piecewise models were investigated and controlled using propensity scores from a logistic regression model designed to predict treatment compliance among eligible at-risk students. The current study is a replication and an extension of previous work in the area. The study is a replication in that a unique new data set was used to investigate the relative efficacy of a summer school initiative. However, by combining the RD design with a relatively well-controlled growth curve analysis, the current investigation was also able to address some of the key design limitations identified in previous summer school evaluations.

METHOD

Data Source

Data analyzed in the present study were obtained from the administrative records of a moderately sized school district in the Pacific Northwest. The district serves close to 6,000 students each year. The student body is approximately 75% White, 14% Latino, 3% African American, 3% Asian American, 3% Native American, and 2% other. Many students face economic challenges. On average, 44% of the students who attend district schools receive a free or reduced price lunch. Recent demographic changes have led to a significant rise in the number of Latino students and non-native English speakers. In the past 10 years, the percentage of Latino students has doubled from 7% to 14%, whereas the number of students classified as English language learners (ELLs) has increased 300%. Although non-native English speakers are still a small fraction of the total student body ($N = 175$; 3%), the district now provides site-based English language instruction to such students at three elementary schools.

Summer Program

As a means of closing the performance gap between strong and struggling readers and ensuring that struggling readers gain the foundational skills requisite to meet reading proficiency targets, the district recently established an academically intensive summer literacy program. The program is offered annually to early elementary school students who did not meet a proficiency cutscore on the spring administration of a grade-appropriate formative literacy assessment. Each year, the program runs over a 5-week period during the middle of the 3-month summer vacation period. Instruction occurs 3.5 hr/day four mornings per week in small class size (<20) environments housed in one central school site. Students receive a minimum of 2 hr of teacher-directed daily reading instruction in the critical beginning reading skills of phonemic awareness (oral blending and segmentation), alphabetic understanding (letter sounds, decoding, phonic analysis), and fluency/automaticity (speed and accuracy in reading connected text). The curricular components of three reading models (Reading Mastery, Open Court, and Horizons) form the basis of instruction. Instruction focuses primarily on basic phonemic, alphabetic, and fluency skills rather than vocabulary and comprehension, as facts and procedural skills are more susceptible to forgetting than conceptual knowledge (Bryk & Raudenbush, 1988; Cooper et al., 1996; Heyns, 1978, 1987), and because the program aims to more closely align with the fundamental "big ideas" and best practices that underlie and facilitate early childhood literacy development (National Reading Panel, 2000).

Daily lessons begin in a whole-group setting in which attendance is taken, homework is checked, and a morning seatwork packet is delivered. Literacy skills are then modeled and practiced in whole- and small-group (3–5 students) settings. Students are grouped based on skill to allow small-group instruction to be maximally supportive and aligned with individual student needs. After a short recess, students cycle through small-group and individual reading activities delivered at mixed-use literacy stations. The literacy stations allow for independent student work and collaborative group learning and also provide an opportunity for need-based differentiated

Early Education and Development 2011.22:649-675.

instruction. Lessons close with a whole-group review period. In each whole- and small-group arrangement, teacher modeling of skills and multiple opportunities for practice with corrective feedback are an explicit part of the curriculum.

Analytic Sample

The study sample was composed of four longitudinally matched first-grade student cohorts ($N = 1,449$). The cohorts consisted of students who entered first grade during the 2003–2004 ($n = 347$), 2004–2005 ($n = 387$), 2005–2006 ($n = 348$), and 2006–2007 ($n = 367$) school years and completed second grade the following year. The overall sample was identified as 48% ($n = 701$) female, 77% ($n = 1,121$) White, 13% ($n = 185$) Latino, 3% ($n = 41$) African American, 3% ($n = 40$) Asian, 3% ($n = 39$) Native American, and 2% ($n = 23$) other. A total of 41% of the sample ($n = 593$) received a free or reduced priced lunch, and 27% ($n = 395$) received Title I services during first grade. A smaller number of students were identified as ELLs (4%, $n = 59$).

Placement and Outcome Measure

Scores on the Test of Oral Reading Fluency (TORF; Children's Educational Services, 1987) were used to determine a student's eligibility for a summer school placement and to measure short- and longer term changes in oral reading fluency. The TORF is a standardized, individually administered test of accuracy and fluency with connected text that has been specifically designed and developed to identify struggling readers and allow for the monitoring of student progress over time. TORF passages and procedures are based on the program of research and development of Curriculum-Based Measurement of Reading (see Shinn, 1989). Student performance is measured by having students read each of three passages aloud for 1 min. Word omissions or substitutions and hesitations of more than 3 s are scored as errors. If a word is self-corrected within a 3-s timeframe, the response is scored as accurate. The median correct words per minute from the three passages indexes the oral reading fluency rate. The test developer reported that test–retest reliabilities obtained from the repeated assessment of elementary students ranged from .92 to .97, whereas alternate-forms reliability of different reading passages drawn from the same level ranged from .89 to .94 (Tindal, Marston, & Deno, 1983). In other studies that have examined the relationship between TORF scores and performance on a variety of reading criterion measures, concurrent validity estimates have ranged from .52 to .91 (see Good & Jefferson, 1998), whereas the relationship between TORF scores and later performance on the third-grade Oregon Statewide

Assessment reading/language subtest has ranged from .67 to .82 (Good, Simmons, & Kame'enui, 2001).

Each academic year, the school district administers the TORF to first- and second-grade students in September, January, and May. First- and second-grade TORF passages are designed to be of equal difficulty within and between grade levels. Students who correctly read fewer than 30 words per minute on the first-grade May assessment are considered to be at heightened risk for future reading difficulty. These students are offered the opportunity to attend the summer enrichment program. In addition, the district has also relied on the professional judgment of teachers and administrators to refer students who score slightly above the 30 words per minute spring benchmark (e.g., a student who reads somewhat fluently but struggles with comprehension may be offered a summer school placement). The 30 words per minute benchmark is used for conceptual and practical reasons. Students producing fewer than 30 words per minute are >20 words per minute below the norm the district uses to define an established reader. Budgetary constraints also allow summer placements to be offered only to students most in need of supplementary instruction.

Across the four cohorts, 17% ($n = 245$) of students in the four first-grade cohorts participated in the summer school program. A total of 66% of participants ($n = 162$) met the cutscore eligibility criterion, whereas 83 students participated on the basis of administrative referral. It should also be noted that 37% of students (98 of 260) who were cutscore eligible declined to participate. Figure 1 displays the relationship between spring TORF scores and the probability of attending summer school. In the figure, it can be seen that the probability of attending summer school declined quite steeply in the vicinity of the eligibility benchmark. However, the "fuzziness" in participation status around the assignment cutoff also serves to illustrate the extent to which self- and administrative selection practices impacted actual program attendance. In the following sections, descriptive statistics are presented on the summer status of students and a series of inferential models are used to account for the differential assignment and selection practices that produced each summer classification.

Predictor Variables

Three summer status indicators were formed in order to estimate and compare the literacy outcomes of program participants; eligible "refusers"; and their nontreated, more reading fluent peers. The first indicator variable distinguished between participants ($n = 245$) and nonparticipants ($n = 1{,}204$). This contrast was used to provide a treatment-on-treated (TOT) estimate of the summer school effect in the first RD model presented here. The

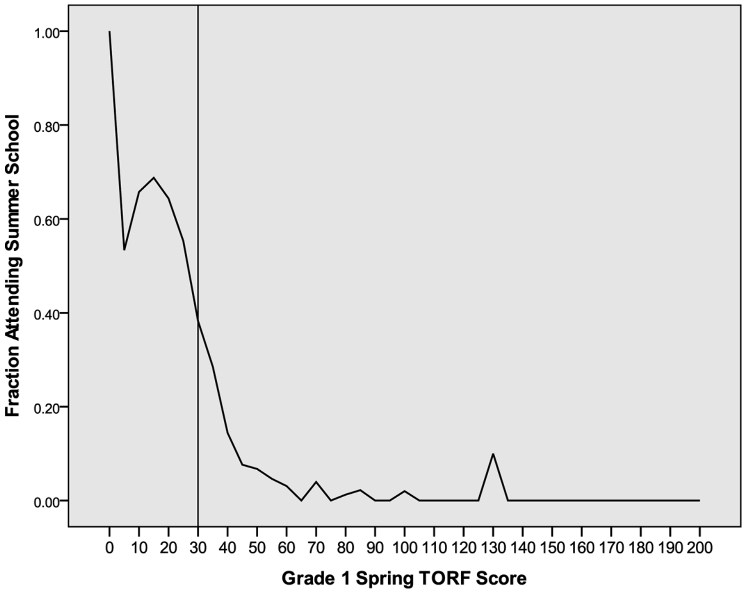

FIGURE 1 Fraction of students attending summer school as a function of first-grade spring Test of Oral Reading Fluency (TORF) score.

second indicator distinguished between cutscore eligible (i.e., eligible participants and eligible refusers, $n = 260$) and cutscore ineligible ($n = 1{,}189$) students whether or not they actually attended summer school. This contrast provided an intent-to-treat (ITT) estimate of the summer school effect in the second RD model presented here. The ITT indicator was also used as an instrumental variable (IV) to take into account the cutscore-based misallocation of students in a third RD model, thereby providing an IV-adjusted TOT estimate of the summer school effect (see Wong, Cook, Barnett, & Jung, 2008).

The final summer status indicator was formed to distinguish between cutscore eligible students who chose to participate in summer school ($n = 162$) and cutscore eligible students who chose not to participate (i.e., refusers; $n = 98$). This indicator was first used as a dependent variable in a logistic regression analysis designed to identify the characteristics of students who were treatment compliant. Along with the resulting propensity score, the third summer status indicator was then used as a predictor in a piecewise growth model designed to estimate the performance of cutscore eligible students prior to, during, and after the close of the summer intervention.

In addition to the summer status eligibility and participation indicators, several demographic characteristics also served as predictor variables.

TABLE 1
Means and Standard Deviations of Study Variables by Summer School Status ($N = 1{,}449$)

	Cutscore Eligible Participants		*Cutscore Eligible Refusers*		*Cutscore Ineligible Participants*		*Cutscore Ineligible Nonparticipants*	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	0.48		0.55		0.43		0.48	
Ethnic minority	0.20		0.19		0.25		0.23	
English language learner	0.04		0.02		0.12		0.04	
Free/reduced lunch	0.51		0.48		0.48		0.38	
Title I student	0.73		0.55		0.52		0.16	
Age at entry to Grade 1	76.40	3.84	77.21	4.52	76.54	3.99	77.25	3.84
Winter TORF Grade 1	10.25	4.69	12.47	5.45	17.70	8.97	46.22	32.15
Spring TORF Grade 1	20.23	5.77	21.83	5.71	43.24	16.91	78.88	32.75
Fall TORF Grade 2	26.47	11.13	23.92	9.67	48.33	17.54	72.68	31.80
n	162		83		98		1,106	

Note. The mean of each dichotomous variable represents the proportion of the sample with the identified characteristic (e.g., 4% of benchmark eligible participants were English language learners). TORF = Test of Oral Reading Fluency.

Indicator codes were used to identify female and ethnic minority students as well as free or reduced price lunch recipients (FRL), ELLs, and students receiving Title I services. Student age at entry to first grade and first-grade winter and spring TORF scores were also used as predictors in one or more analyses. Descriptive statistics for the study variables by summer school status are shown in Table 1. As expected, the statistics in Table 1 demonstrate that cutscore eligible students scored lower on the winter and spring TORF assessments than their cutscore ineligible peers. On average, the receipt of Title I services was also higher among cutscore eligible students. For cutscore eligible students (i.e., summer school participants and refusers), winter and spring TORF scores and the distribution of student characteristics were relatively similar (with the exception of Title I status). However, among cutscore ineligible students, winter and spring TORF scores were notably higher, whereas the proportions of ELL, FRL, and Title I students were lower in the group of nonparticipants.

Analytic Procedures

A series of data analyses were conducted to model selection and estimate program effects on the summer and academic year learning outcomes of students. To identify the effect of the summer school intervention, a series of single-level RD models was estimated. RD models are particularly appropriate when a treatment or program is allocated to those in need using a cutscore-based assignment protocol (Shadish et al., 2002; Trochim, 1984; Wong et al., 2008). Model estimates derived from RD-based designs yield inferences with strength comparable to those associated with randomized experiments, as the selection mechanism for treatment assignment (a cutscore) is completely known and controlled by the researcher or program administrator (Campbell, 1969; Thistlethwaite & Campbell, 1960). In cases when the assignment protocol is strictly followed, RD designs are not susceptible to traditional selection bias concerns or most other internal validity threats. RD designs are also particularly useful in applied treatment contexts, as they are consistent with how many policies and programs are implemented (Shadish et al., 2002; Trochim, 1984).

Similar to a randomized experiment, however, the strength of the inference available from an RD design is weakened to the extent that those assigned to treatment fail to comply and/or when treatment is delivered to members of the control group. In an RD context, misallocation occurs when cutscore eligible individuals refuse treatment or when treatment is delivered to cutscore ineligible individuals (i.e., those with scores on the assignment variable that fall outside of the designated treatment range). Under these circumstances, a sharp RD design becomes fuzzy and the analysis must adjust for the potential bias that arises from the nonrandom selection of individuals into conditions. In addition, the validity of an RD inference is also contingent upon correct modeling of the functional form of the relationship between the assignment and outcome variables (Imbens & Lemieux, 2008). Identification of the functional form involves specification of higher order polynomials and associated interaction terms in the statistical model. Inclusion of these terms enables the analyst to determine whether a nonlinear relationship or interaction between the assignment and treatment variable is responsible for an observed discontinuity (Shadish et al., 2002; Trochim, 1984; Wong et al., 2008).

To examine the effects of selection and to take into account the fuzzy RD design that emerged in the current study, estimates from a series of RD models are presented here. The first RD model provides a TOT estimate of the summer school effect for those students who participated in the summer program regardless of assignment status, whereas the second model yields a more conservative ITT estimate for those who were assigned to treatment whether they participated or not. The third RD model provides an IV-adjusted TOT estimate in which assignment status was used as an instrument to account for potential bias in the nonrandom take-up and refusal of treatment. In all models, the outcome was the change in literacy performance between the final assessment in first grade and the initial assessment in second grade. Key predictors were the first two summer school status indicators described previously and the test score assignment variable centered on the cutscore (i.e., first-grade spring TORF = 30). By virtue of the centering used to represent the assignment variable, the coefficient associated with each treatment status indicator is an estimate of the local average treatment effect at the cutscore, the point at which the treatment and control groups were most similar prior to the onset of the intervention (see Shadish et al., 2002; Trochim, 1984).

Equation 1 presents the basic form of the RD model, in which Y_i is the summer change in literacy status for student *i*, β_0 is the intercept, β_1 is the effect of treatment, β_2 represents the relationship between the assignment score and the outcome, β_3 represents the treatment status by the centered assignment score interaction, and e_i is a residual term.

$$\begin{aligned} Y_i = \beta_0 + \beta_1(\textit{Treatment Status}) + \beta_2(\textit{Assignment Score} - \textit{Assignment Cutoff}) \\ + \beta_3(\textit{Treatment Status} \times \textit{Centered Assignment Score}) + e_i \end{aligned} \tag{1}$$

As a means of identifying potential nonlinear relationships between the assignment variable and the outcome, quadratic and cubic representations of the assignment variable and associated interaction terms between treatment status and the assignment variable were also added to the base model. The inclusion of these terms enabled estimation of the correct functional form of the regression, thereby protecting against the identification of a spurious discontinuity at the cutscore for treatment assignment.

In a subsequent set of analyses, piecewise growth models were applied to the TORF scores of cutscore eligible students to estimate performance prior to, during, and after the summer intervention. Two-level unconditional and conditional piecewise growth models were estimated using HLM Version 6.0 (Raudenbush, Bryk, Cheong, & Congdon, 2004). Two-level piecewise growth models (observations within students) were estimated, as students changed academic year classrooms each year and eligible nonparticipants (i.e., refusers) did not experience a nested classroom structure in the summer. The unconditional model provided estimates of the mean literacy status and growth of students and the amount of student-to-student variation across each growth trajectory component. The conditional model allowed for the examination of relationships between student predictor variables and each of the student literacy outcomes estimated in the unconditional model (Raudenbush & Bryk, 2002). Of particular interest was estimation of the time-specific (i.e., piecewise) performance of cutscore eligible students who participated in summer school and their cutscore eligible peers who declined to participate. Equation 2 specifies the basic form of the unconditional piecewise growth model. In Equation 2, it can be seen that student literacy performance is conceived as a function of one status and two growth parameters, one for each of the time frames of interest.

$$Y_{ti} = \pi_{0i} + \pi_{1i}(Summer\ Break) + \pi_{2i}(Second\ Grade) + e_{ti} \tag{2}$$
$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$
$$\pi_{2i} = \beta_{20} + r_{2i}$$

In the piecewise model, time was represented both as a continuous variable capturing the linear change in student literacy across multiple assessments during the second-grade academic year and as a dichotomous variable (coded 0, 1) capturing the difference in literacy performance from the final literacy assessment in first grade to the initial literacy assessment in second grade. The coding used to represent the change in literacy across the different time periods defined the status parameter (π_{0i}) as the expected performance of student *i* at the end of first grade. After the parameters of the unconditional model were estimated, variation in the literacy status (π_{0i}) and growth rates (π_{1i}, π_{2i}) of students was modeled conditionally as a function of the third summer status indicator variable described previously and a propensity score resulting from a logistic regression analysis predicting summer school participation among cutscore eligible students. The propensity score served as a partial control for possible selection effects in the estimation of the short- and longer term performance of cutscore eligible participants and nonparticipants.

RESULTS

RD Models

Results associated with the RD models are presented in Table 2. In the first TOT model, which contrasted participants with nonparticipants regardless of eligibility status, the local average treatment effect was estimated as a statistically significant 5.7 words per minute change in oral reading fluency over the summer vacation period. Combined with the coefficient representing the intercept, treated students who scored at the cutscore for treatment assignment were estimated to experience a 7 words per minute gain over the summer (i.e., $5.69 + 1.33 = 7.02$). Translated into an effect size, the 5.7 words per minute discontinuity represents a 0.40 *SD* difference between

TABLE 2
Treatment-on-Treated (TOT) and Intent-to-Treat (ITT) Regression Discontinuity Model Results

	TOT		*ITT*		*IV-Adjusted TOT*	
Variable	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	1.33	0.74	5.66	1.13***	1.80	1.08
Spring TORF	−0.21	0.05***	−0.44	0.07***	−0.24	0.06***
Participant	5.69	1.29***			5.80	1.30***
Participant × TORF	0.25	0.08**			0.24	0.09**
Cutscore Eligible			0.45	3.17		
Eligible × TORF			0.28	0.96		

Note. Dummy codes were used to identify program participants in the TOT models and benchmark eligible students in the ITT model. IV = instrumental variable; TORF = Test of Oral Reading Fluency.
$^{**}p < .01$. $^{***}p < .001$.

Early Education and Development 2011.22:649-675.

participants and nonparticipants at the assignment cutscore. In addition, although none of the higher order terms and associated interactions were statistically related to the outcome, a statistically significant linear interaction between treatment status and the assignment variable was observed. The positive coefficient indicates that the slope relating spring TORF scores to the change in summer oral reading fluency was positive for students scoring below the assignment cutscore but negative for students scoring above the cut. This result suggests that summer losses in oral reading fluency were predicted to be greatest in the lowest and highest scoring students, with the largest summer losses projected among the highest performers on the spring TORF assessment.[1]

In the second RD model, which contrasted benchmark eligible students and their benchmark ineligible peers regardless of their actual participation status (the ITT model), the intercept indicates that ineligible students, including ineligible students who received treatment, had an average change in TORF of 5.7 at the cutpoint. However, the effect of eligibility at the assignment cutoff was estimated as a statistically nonsignificant 0.45 words per minute change in performance over the summer, showing that eligibility in and of itself was not responsible for summer gains.

[1]One might question whether some of the observed results could be due to the use of a gain score. The same models presented here were also run using the second-grade fall TORF score as the outcome. Results showed a statistically significant positive relationship between spring first-grade and fall second-grade TORF scores, but coefficients in the regression models were essentially identical to those reported in the text.

Estimates associated with the third RD model are the final set of results presented in Table 2. The IV-adjusted TOT estimates differed from those associated with the ITT model but were similar to those associated with the unadjusted TOT model.[2] The local average treatment effect for program participants was estimated as a statistically significant 5.8 words per minute change in oral reading fluency over the summer vacation period ($d = .40$) for a total summer gain of 7.6 words (i.e., $5.80 + 1.80 = 7.60$). In this model, spring TORF scores again interacted with treatment status. The differential relationship between spring TORF performance and the change in summer reading fluency and the resulting discontinuity between treatment groups can be seen in the displacement of the separate smoothed regression functions displayed in Figure 2. To the left of the reference line, it can be seen that higher scoring cutscore eligible students were estimated to have greater average summer gains than their lower scoring cutscore eligible peers. For cutscore ineligible students to the right of the reference line, oral reading fluency gains were only estimated for those with spring TORF scores closest to the cutscore for treatment assignment. Summer oral reading fluency losses were predicted for the reminder of cutscore ineligible students. Further consideration of the size and direction of the within-group slopes is provided in the Discussion section.[3] In all, 18% of the variation in summer reading performance was accounted for using the set of predictor variables ($p < .001$).

Logistic Regression Model

Prior to estimating the relationship between treatment status and the summer and second-grade change in oral reading fluency of benchmark eligible students, a logistic regression model was estimated to predict treatment compliance status ($0 = \text{noncompliant}$, $1 = \text{compliant}$). The logistic model revealed a statistically significant relationship between the set of predictors and the grouping variable, $\chi^2(7,\ N = 260) = 23.49$, $R^2 = .118$, $p < .001$.

[2]In an RD model that included demographic covariates, a 6.7 words per minute discontinuity between participants and nonparticipants was observed ($d = .47$). Student age was positively related and Title I status was negatively related (non–Title I advantage) to summer changes in oral reading fluency ($p < .05$). The block of covariates accounted for an additional 2% of the variation in the outcome.

[3]As a further check of the functional form of the regression, the sample was limited to those with scores of 50 or less on the assignment variable (first-grade spring TORF). In this limited sample, the interaction between treatment status and the assignment score was not statistically significant. Instead, a statistically significant negative linear relationship between the assignment score and summer reading gains was observed. Nonetheless, the estimated treatment effect in the reduced sample was nearly identical to that observed in the full sample. The IV-adjusted TOT estimate revealed a statistically significant 5.6 words per minute discontinuity at the cutscore.

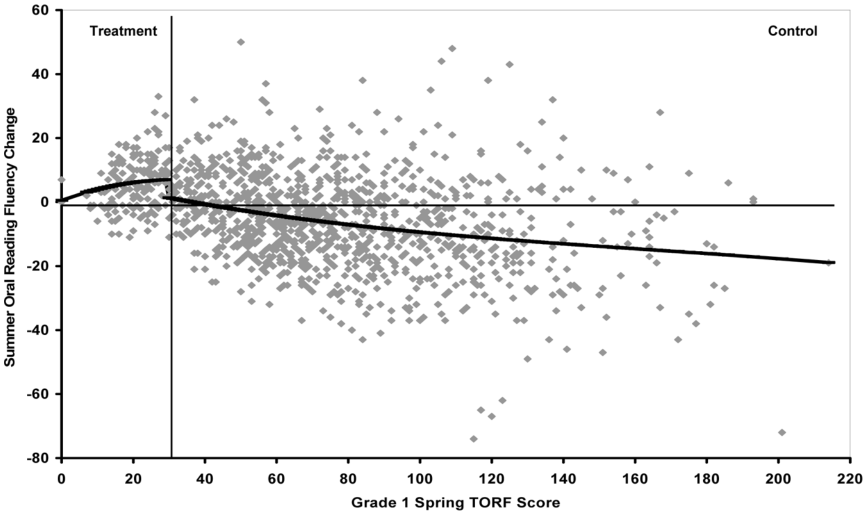

FIGURE 2 Summer oral reading fluency change as a function of spring Test of Oral Reading Fluency (TORF) score and summer treatment status.

Coefficients, standard errors, Wald tests, and odds ratios are presented in Table 3. It can be seen in Table 3 that Title I students had statistically higher predicted odds of participating in summer school relative to their non-Title I peers. The odds of treatment compliance for Title students were 49% higher than their non-Title I peers. Also statistically related to treatment compliance was winter of first grade TORF performance. The negative coefficient associated with the winter TORF variable indicates that for each additional word per minute correctly read during the winter of Grade 1, the odds of compliance were 8% lower. Overall, the logistic model correctly classified the compliance status of 65% of cutscore eligible students. Associated propensity scores (i.e., predicted compliance probabilities) were retained and used as an adjustment factor in the conditional growth model presented here.

Unconditional Piecewise Growth Model

Estimation of the unconditional model revealed that cutscore eligible students correctly read an average of 21 words per minute on the spring assessment in first grade ($\beta_{00} = 20.68$), whereas over the summer between first and second grade, average reading fluency increased by slightly less than 4 words per minute ($\beta_{10} = 3.96$). During second grade, cutscore eligible students'

TABLE 3
Logistic Regression Model Predicting Treatment Compliance Among Cutscore Eligible Students

Variable	*b*	*SE*	*Wald*	*Exp(b)*
Female	−0.188	0.272	0.478	0.793
Minority	−0.082	0.361	0.051	0.915
English language learner	1.292	1.057	1.493	1.275
Free/reduced lunch	−0.089	0.285	0.098	0.907
Title I student	0.719	0.292	6.077	1.487*
Age at entry to Grade 1	−0.059	0.033	3.166	0.943
Winter TORF Grade 1	−0.089	0.029	9.478	0.915**

Note. A dummy code was used to identify program participants. TORF = Test of Oral Reading Fluency.
$^{*}p < .05$. $^{**}p < .01$.

average reading fluency increased at a rate of 23.5 words per assessment occasion ($\beta_{20} = 23.57$). Associated variance component estimates revealed that, with the exception of the variability in the first-grade exit status parameter, the variation observed in all other growth model parameters was statistically significant ($p < .05$). This suggests that students differed in the rate of acquisition of oral reading fluency during the summer and second grade but that sampling variability remained an explanation for student-to-student differences in end of first-grade status. Although sampling error remains a viable explanation for the observed differences in TORF first-grade exit status, interest in examining the pre-summer TORF performance of cutscore eligible students who accepted or declined a summer school placement led to retention of the exit status parameter as a random coefficient in the conditional model presented here.

Conditional Piecewise Growth Model

Table 4 presents the results associated with the conditional model. Predictors are separated by row, outcomes by column. Coefficients associated with the treatment status contrast are presented in the second row of Table 4. The treatment status estimates indicate that cutscore eligible summer school participants ended first grade with the ability to read the same number of words as their cutscore eligible peers who declined to participate in the summer program. Over the summer between first and second grade, cutscore eligible refusers maintained their level of oral reading fluency, whereas cutscore eligible program participants gained an average of 6 words per minute (i.e., $0.56 + 5.51$), allowing program participants to enter second

TABLE 4
Conditional Piecewise Growth Model Predicting the Status and Growth in Oral Reading Fluency

Variable	*First Grade TORF Exit Status*	*Summer TORF Change*	*Second Grade TORF Growth*
Intercept	20.74 (0.59)***	0.56 (0.77)	25.13 (0.39)***
Participant	−0.08 (0.73)	5.51 (1.09)***	−2.48 (1.17)*
Compliance probability	−0.15 (0.02)***	−0.14 (0.04)**	−0.01 (0.04)

Note. A dummy code was used to identify program participants. Standard errors are in parentheses. TORF = Test of Oral Reading Fluency.
*$p < .05$. **$p < .01$. ***$p < .001$.

grade with a higher rate of oral reading fluency. However, during second grade, cutscore eligible summer participants and refusers grew at a divergent rate (22.65 and 25.13 words per assessment, respectively). The 2.48 words per assessment advantage for cutscore eligible refusers was statistically significant ($p < .05$) and of enough strength to allow for a closing of the gap in reading fluency that emerged over the summer. These outcomes were obtained after student differences in the propensity to participate in the

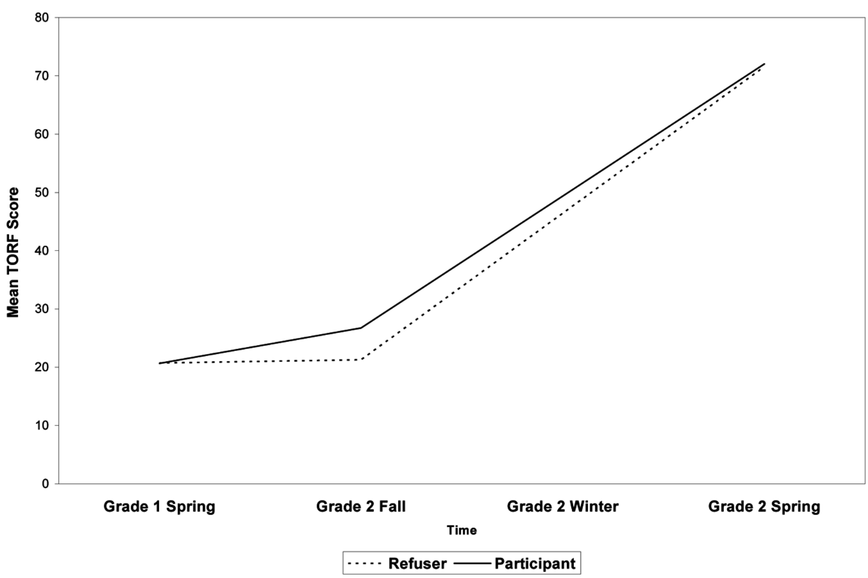

FIGURE 3 Mean Test of Oral Reading Fluency (TORF) score as a function of summer treatment status and time among cutscore eligible students.

summer program were accounted for.[4] Figure 3 presents the trajectory of oral reading fluency growth for cutscore eligible participants and refusers over the summer between first and second grade as well as over the second-grade school year.

DISCUSSION

The purpose of the current study was to conduct a series of analyses on data from four first-grade student cohorts as a means of evaluating selection and efficacy issues surrounding an academically intensive summer literacy intervention. Drawing on a cutscore treatment assignment protocol, the change in summer literacy performance for participants, refusers, and higher scoring cutscore ineligible students was estimated. The short- and longer term oral reading fluency outcomes for students in both cutscore eligible categories (i.e., participants and refusers) were also examined. RD model estimates revealed that students scoring at the cutscore for admission had summer literacy outcomes that varied with respect to their participation status. More specifically, relative to a nonparticipant peer who scored just above the eligibility benchmark, a participating student who scored just below the cut for program admission was estimated to gain 5.8 more words per minute in reading fluency over the summer months. In addition, the use of piecewise growth models to examine the short- and longer term performance of cutscore eligible students revealed that participants in the summer program increased their absolute and relative levels of oral reading fluency over the summer. However, summer school participants' rate of reading fluency increased at a slower pace than that of their noncompliant peers during the following school year. Together these results suggest that supplemental summer instruction delivered to at-risk students may promote literacy gains during the otherwise challenging summer months and thereby serve as a useful intermediary tool for K–12 stakeholders seeking to keep struggling readers on track toward proficiency.

The results of the current study were generally consistent with previously published findings regarding the summer achievement performance of

[4]In a piecewise growth model that specified individual demographic covariates instead of propensity scores, the pattern of treatment group similarities and differences was analogous to that of the propensity adjusted model. Benchmark eligible participants were not statistically different from nonparticipants at the end of first grade ($p > .05$), gained 5.51 more words per minute over the summer, and grew at a 2.51 words per assessment deficit in Grade 2 ($p < .05$). First-grade winter TORF was positively related to first-grade exit status and summer oral reading fluency gain, whereas FRL status (non-FRL advantage) and ELL status (non-ELL advantage) were related to exit status and summer change, respectively ($p < .05$).

students who participate in a supplementary enrichment program. As reported in other investigations (see Borman et al., 2005; Borman & Dowling, 2006; Cooper et al., 2000; Matsudaira, 2008; Schacter & Jo, 2005), supplemental summer instruction delivered to struggling readers not only prevented a summer learning loss but also facilitated a summer literacy gain. However, it should be noted that the summer setback experienced by the nontreated, more reading fluent students in this study was somewhat incongruent with the summer performance of advantaged students reported in other investigations of summer learning outcomes (Cooper et al., 1996). Although higher achieving students tend to maintain or slightly increase their performance levels across the summer months, stronger readers in the current study experienced a change in summer literacy more akin to that observed among struggling readers or disadvantaged students in other studies. This discrepancy is likely due at least in part to the type and timing of assessments that were used in the current study. Reading fluency measures can be more sensitive to practice effects than the achievement measures commonly used in other studies of summer learning outcomes. For students able to read a large number of words per minute at the end of one school year, a fluency assessment at the beginning of the following school year may thus reveal a noticeable drop in the number of words per minute read unless active reading engagement was a common feature of the summer vacation period. However, strong readers with an observed summer drop in fluency tend to quickly regain their previous level of performance once reengaged in the scholastic environment (S. K. Baker, personal communication, November 17, 2009).

The general pattern of summer reading fluency outcomes observed in the current study may also be due in part to the unique nature of the sample under investigation. District students receive a high-quality reading program with an emphasis on fluency development. The instructional experience enables many students who otherwise come from challenging demographic backgrounds to perform above "average" when engaged during the academic year. However, during the summer, some of these higher performing students do not have access to the same level of educational resources and therefore experience the summer slide that is commonly observed among economically disadvantaged youth. Currently, it is not known whether a similar result would also be observed on a comprehension or vocabulary measure or in another content domain, but it is likely that the summer performance of the "above average" readers in this sample would be more similar to that reported in other investigations of disadvantaged youth (see Alexander et al., 2001; Bryk & Raudenbush, 1988). Data on youths' summer reading engagement and performance on other reading measures would clarify whether the decline in the oral reading fluency of some relatively successful academic year readers may be attributed in part to a local regression to the mean artifact and/or less consistent skill maintenance and building (i.e., reading practice) during the summer (Burkam, Ready, Lee, & LoGerfo, 2004; Kim, 2004).

As results from the current study offer additional insight into the potential impact of a summer literacy intervention as well as the summer learning patterns of youth, it is necessary to further consider the context in which the findings originated. As with many other investigations of summer school programs, students (by design) were offered supplementary instruction on the basis of need, and student participation was voluntary. Potential inferential issues associated with the cutscore treatment protocol were addressed in part through the estimation of alternatively specified RD models. Resulting differences between the TOT and ITT estimates provided a useful demonstration of how inferences regarding program performance can change when administrative and self-selection processes combine to form treatment classifications distinct from those originally intended. In particular, the ITT estimate indicated that at the cutscore for program admission there was no statistically significant difference in performance between eligible and ineligible students. In contrast, the TOT and IV-adjusted TOT estimates revealed a program participation effect. Taken together, the RD estimates suggest that summer literacy enrichment is likely beneficial for those who receive it (i.e., instructional exposure matters), but without careful control over the allocation of summer placements and the voluntary self-selection into treatment, not only is the distribution of resources not optimized but estimation of program effects may be blunted.

In addition to concerns related to the fuzziness of the assignment cutscore, collection of data on the fidelity with which the literacy intervention was delivered and data on the summer activities of students would also serve to further enrich experts' understanding of the costs and benefits associated with offering summer instruction. Data on teaching practice and classroom process would aid in documenting the range of instructional variation experienced by student participants, whereas additional data on the availability and use of home-, community-, and school-based educational resources would better contextualize the summer literacy growth patterns of all students. Some of these data may also be helpful in further accounting for the differential participation rates and learning trajectories of cutscore eligible students. For example, despite relatively similar demographic profiles and comparable end-of-first-grade reading fluency scores, 37% of students who were offered a summer placement declined to participate. Whether differential administrative encouragement, parental perceptions and beliefs, student self-efficacy, or a combination of these and other unobserved factors dissuaded some students (and their parents) from accepting a summer school placement, this decision ultimately resulted in a very modest half of 1 word gain in oral reading fluency over the summer months. However, by acquiring oral reading fluency skills at a faster rate during the following school year, summer school refusers were able to close the gap that had opened over the previous summer with cutscore eligible participants. Currently, it is not clear if some of the same environmental conditions and personal attributes that influenced the decision to opt out of summer school also enabled these students to recover when reengaged in the scholastic environment or whether focused instruction and heightened attention from school staff in reaction to student nonparticipation provided the requisite scaffolding the students needed to overcome their summer stagnation.

Despite the need for continued research on student and parental perceptions and beliefs, as well as more fine-grained examination of the educational experiences of youth, the demonstration that student literacy development varies with respect to the receipt of supplemental summer instruction highlights the potential of using the summer vacation period as an avenue for addressing NCLB accountability pressures. Moreover, the provision of targeted summer reading instruction may hold a further two-fold appeal for K–12 stakeholders seeking to provide or engender an effective and equitable scholastic experience for students. First, given the observation that reading trajectories establish early, become increasingly discrepant, and are often resistant to change (Juel, 1988; Schatschneider, Francis, Carlson, Hetcher, & Foorman, 2004; Stanovich, 1986), early and active intervention efforts in beginning reading can be viewed as a primary way of addressing initial reading difficulties and preventing them from snowballing into long-term reading failure (Coyne & Harn, 2006; Torgesen, 2002). A second appeal of offering summer reading instruction is the ease with which educational leaders can describe program logic to stakeholders. In particular, educational leaders can easily convey and a range of K–12 stakeholders can readily understand that as a result of the school offering academically focused summer contact hours, students have a clear opportunity to benefit from an extended flow of educational resources. Further traction can be gained when a contrast is drawn with the common status quo of the "no summer treatment" alternative (see Downey et al., 2004, 2008). Therefore, coupling the straightforward theory of change that underlies the initiative (i.e., keeping the educational resource faucet fully open) with presentation of the well-documented shift in summer learning trajectories enables educational leaders to effectively describe how a targeted summer school program can keep at-risk students on track toward proficiency.

The literature on summer learning rates and supplementary summer instruction shines light on the challenge and opportunities the summer vacation period presents to students and K–12 stakeholders. This literature also highlights a number of challenging methodological issues that can undermine the validity of inferences drawn from the study of summer learning outcomes. The range of methodological issues identified in the literature on summer learning and associated school-based interventions suggests that additional study of summer achievement outcomes with diverse samples, better controls, appropriately scaled measures, and thorough documentation of the fidelity with which an intervention is delivered to students is requisite to further elucidate the manner and degree to which supplemental summer instruction can serve to offset the summer slide commonly experienced by disadvantaged or academically at-risk students. Data on the nonprogram summer activities of students are also requisite to allow for a more nuanced examination of the manner in which the summer experiences of comparison or control students differ from those of the treatment group. For those who receive or decline the instructional treatment, additional information on the degree of access to and utilization of educational resources in the home and community and the reasons why they decline to participate can provide for greater understanding of their summer achievement outcomes. Likewise, the availability of data on instructional practice and process is useful in distinguishing the specific and perhaps important ways in which instruction varies across classrooms and treatment sites (Borman, 2000; Borman et al., 2005; Borman & Dowling, 2006; Cooper et al., 2000). Many of these design issues are applicable to the proper evaluation of any field-based intervention, but for educational leaders faced with a limited budget, great student need, and the challenge of meeting NCLB annual yearly progress objectives, additional rigorous examination of a popular yet understudied approach to promoting student success is urged to ascertain the degree to which the findings from the current study will generalize more broadly.

ACKNOWLEDGMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090369 to the University of Oregon. The opinions expressed are those of the authors and do not represent the views of the Institute of Education Sciences or the U.S. Department of Education.

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.

Austin, G. R., Roger, B. G., & Walbesser, H. H. (1972). The effectiveness of summer compensatory education: A review of the research. *Review of Educational Research*, *42*, 171–181.

Borman, G. D. (2000). The effects of summer school: Questions answered, questions raised. *Monographs of the Society for Research in Child Development*, *65*(1), 119–127.

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

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.

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. (2004). Is the school calendar dated? Education, economics, and the politics of time. In G. D. Borman & M. Boulay (Eds.), *Summer learning: Research, policies, and programs*, (pp. 3–23). Mahwah, NJ: Erlbaum.

Cooper, H., Charlton, K., Valentine, J. C., & 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.

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), 1–118.

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.

Cotton, K. (2001). *New small learning communities: Findings from recent literature*. Portland, OR: Northwest Regional Educational Laboratory.

Coyne, M. D., & Harn, B. A. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools*, *43*(1), 33–43.

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.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*. Boulder, CO: Westview Press.

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, NY: Guilford Press.

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*(3), 257–288.

Heyns, B. (1978). *Summer learning and the effects of schooling*. New York, NY: Academic Press.
Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development*, *58*, 1151–1160.
Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*, 615–635.
Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, *86*(1), 226–244.
Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437–447.
Kim, J. (2004). Summer reading and the ethnic achievement gap. *Journal of Education for Students Placed at Risk*, *9*(2), 169–188.
Lee, V. E., Burkam, D. T., Ready, D., Honigman, J., & Meisels, S. J. (2006). Full-day versus half-day kindergarten: In which program do children learn more? *American Journal of Education*, *112*, 163–208.
Lee, V. E., & Smith, J. B. (2001). *Restructuring high schools for equity and excellence: What works*. New York, NY: Teachers College Press.
Matsudaira, J. D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics*, *142*, 829–850.
McMillen, B. J. (2001). A statewide evaluation of academic achievement in year-round schools. *Journal of Educational Research*, *95*(2), 67–74.
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* (National Institutes of Health Publication No. 00-4754). Rockville, MD: National Institutes of Health.
No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* (2nd ed.). Thousand Oaks, CA.
Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.
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.
Schatschneider, C., Francis, D. J., Carlson, C. D., Hetcher, J. M., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*(2), 265–282.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
Shinn, M. (1989). *Curriculum-based measurement: Assessing special children*. New York, NY: Guilford Press.
Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–406.
Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology*, *51*, 309–317.
Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Rep. No. 109). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.
Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, *40*, 7–26.
Trochim, W. M. K. (1984). *Research design for program evaluation: The regression discontinuity approach*. Newbury Park, CA: Sage.

von Hippel, P. T. (2007, August). *What happens to summer learning in year-round schools?* Paper presented at the annual meeting of the American Sociological Association, New York City, NY.

Walston, J. T., & West, J. (2004). *Full-day and half-day kindergarten in the United States: Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99* (National Center for Education Statistics Publication No. 2004-078). Washington, DC: U.S. Government Printing Office.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*(1), 122–154.