Early Childhood Research Quarterly 28 (2013) 24–32

Contents lists available at SciVerse ScienceDirect

Early Childhood Research Quarterly

Summer school effects in a randomized field trial☆

Keith Zvoch*, Joseph J. Stevens

Department of Educational Methodology, Policy, and Leadership, College of Education, 5267 University of Oregon, Eugene, OR 97403-5267, United States

ARTICLE INFO

Article history:
Received 2 March 2011
Received in revised form 26 April 2012
Accepted 14 May 2012

Keywords:
Summer school
Summer learning
Randomized field trial

ABSTRACT

This field-based randomized trial examined the effect of assignment to and participation in summer school for two moderately at-risk samples of struggling readers. Application of multiple regression models to difference scores capturing the change in summer reading fluency revealed that kindergarten students randomly assigned to summer school outperformed their control group peers by .60 of a standard deviation in an intent-to-treat analysis. For the first grade sample, the intent-to-treat estimate was over three quarters of a standard deviation. The contrast in performance was greater when the comparison was focused more specifically on the change in literacy between treatment participants (i.e., randomly assigned students who actually attended summer school) and students randomly assigned to the control group and in analyses that explicitly adjusted for non-compliance with treatment assignment. These results support the experiential intuition of school district personnel regarding the benefits of summer school and suggest that targeted summer instruction can be a useful strategy to support student learning over the summer months.

Examination of the annual pattern of achievement growth reveals that student learning trajectories are largely positive during the academic year but flatten or become negative during the summer months (Alexander, Entwisle, & Olson, 2001; Downey, von Hippel, & Broh, 2004; Downey, von Hippel, & Hughes, 2008; Raudenbush, 2004). The commonly observed slide or drop in learning that occurs during the summer provides an additional unwelcome burden for school district personnel charged with reducing achievement gaps and ensuring that all children reach grade level proficiency as required by the No Child Left Behind (NCLB, 2002) federal legislation. With responsibility for the learning that occurs both within and outside of the traditional nine-month academic year, some school districts have moved to offset summer learning losses by either adopting a year-round schedule for all students (Cooper, Valentine, Charlton, & Melson, 2003; von Hippel, 2007) or implementing targeted summer programs for those students most at-risk of poor summer outcomes (McCombs et al., 2011).

The adoption and implementation of a targeted summer program has become a popular choice as districts are able to focus limited resources on students (e.g., struggling readers) who are in greater need of strategic supplemental instruction. However, the need-based delivery of summer instruction to students struggling to acquire academic skills and knowledge is a challenge in the estimation of the causal impact of programmatic attempts at stemming the summer slide. For researchers and evaluators, the inability to clearly separate instructional effects from the distinct background characteristics of summer school participants and their peers has served to weaken the inferences drawn from various field-based studies (see Borman & Dowling, 2006; Cooper, Charlton, Valentine, & Muhlenbruck, 2000; Zvoch & Stevens, 2011). In the current paper, the difficulty associated with comparing non-equivalent student groups is addressed through an investigation that examines the relative efficacy of one summer school program using a unique source of data obtained in conjunction with a randomized field trial.

The pattern of achievement growth established in the analysis of data obtained from local and nationally representative sources (i.e., positive academic year learning and slowing or negative achievement growth over the summer vacation period) demonstrates the benefit of providing school-based instruction to children (Alexander et al., 2001; Downey et al., 2004, 2008; von Hippel, 2009; Zvoch, 2009). The literature on academic learning trajectories also indicates that while students from divergent backgrounds typically acquire subject knowledge at a similar rate during the school year, summer learning outcomes are more apt to vary by content area and poverty level. In particular, summer mathematics losses tend to be greater than reading losses and interactions with student socio-economic status (SES) are common. Whereas high-SES students tend to make modest gains in reading and mathematics over

☆ 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.

* Corresponding author.
E-mail addresses: kzvoch@uoregon.edu (K. Zvoch), stevensj@uoregon.edu (J.J. Stevens).

0885-2006/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ecresq.2012.05.002

the summer, middle-SES students maintain or slightly improve on pre-summer reading levels, but lose ground in mathematics. For low-SES students, losses in reading and mathematics are the norm (see Burkam, Ready, Lee, & LoGerfo, 2004; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; McCombs et al., 2011). In all, when averaged across content area and poverty level, student scores after summer are approximately a tenth of a standard deviation lower than before the beginning of the summer vacation period (Cooper et al., 1996).

Explanations for the content area summer performance differences are based on the frequency with which reading and mathematics skills can be practiced outside of the scholastic environment. As home and community environments generally provide more opportunity for continued development of reading literacy (e.g., books in the home, libraries, educational media) than for regular practice of the factual and procedural skills that underlie elementary and middle school mathematics, children are better able to maintain or increase proficiency in reading than in mathematics (Bryk & Raudenbush, 1988; Cooper et al., 1996). An extension of the same opportunity-based logic is also used to account for the differential summer learning of students from divergent economic backgrounds. Proponents of the "faucet theory" argue that the summer slide in reading and mathematics achievement commonly experienced by disadvantaged students follows from the restricted flow of educational resources during the summer months (Alexander et al., 2001; Entwisle, Alexander, & Olson, 1997). Unlike their more-advantaged peers who have access to a variety of educational materials in the home as well as support for fee-based educational enrichment programs, disadvantaged students' summer environments tend to offer less psychological support (e.g., lower educational expectations) and fewer of the unstructured or structured activities (reading books, library visits) that are conducive to maintaining and building academic skills (Benson & Borman, 2010; Entwisle et al., 1997; Heyns, 1978, 1987).

Given the body of literature that demonstrates the challenges faced by students during the summer vacation period, educational leaders have increasingly turned toward the provision of supplemental summer instruction to keep their most at-risk students (i.e., low performing and/or economically disadvantaged students) tracking toward proficiency. "Summer school" is seen as a viable option for districts facing NCLB accountability pressures as instruction can be strategically delivered (McCombs et al., 2011) and state or federal Title I resources can frequently be used to fund the program (Fairchild, Smink, & Stewart, 2009). Relative to the attendant disruption and questionable efficacy associated with a school-wide change to a year round schedule (Cooper et al., 2003; von Hippel, 2007), targeted summer school programs are also more easily tolerated and supported by parents and other system stakeholders (American After 3PM, 2010). In recent years, mandatory summer programs for low-achieving students have been implemented in large urban school districts and smaller suburban districts alike. Voluntary summer programs are also now in use in a variety of districts and schools throughout the United States (McCombs et al., 2011). 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, with higher participation rates among ethnic-minority and low-income students (American After 3PM, 2010). The summer school programs that children are mandated or encouraged to attend vary in terms of length of instruction as well as focus and content (e.g., remediation vs. acceleration, academics vs. recreation), but many are now designed to provide academically rigorous instruction over a several-week period to those most at-risk of poor summer learning outcomes (Cooper et al., 2000; McCombs et al., 2011).

Despite the non-trivial percentage of students that participate in a summer school program, the evidential basis supporting the efficacy of supplemental summer instruction is somewhat weak. A comprehensive review of the summer school literature outlined a number of methodological shortcomings that have limited the strength of inferences associated with many reported investigations (Cooper et al., 2000). In this review, the common use of one group pre-post and nonequivalent control group designs was noted as a chief concern as was the need-based selection and voluntary participation of students. Yet, when meticulously summarized over diverse samples, research designs, instructional conditions, and program foci, Cooper and colleagues concluded that summer instructional support enabled disadvantaged students to maintain or slightly improve upon their academic year reading and mathematics gains while instruction delivered to middle-class students served to promote continued acquisition of reading and mathematics skills. Cooper et al. (2000) also concluded that early-elementary and late-secondary students tended to derive the most benefit from summer school participation. The overall size of the summer school effect across all studies and samples reviewed was approximately a quarter of a standard deviation. However, the average effect for middle-class students ($d = \sim.50$) was larger than for lower income students ($d = \sim.20$). More recent studies that employed randomized designs have also provided suggestive evidence of positive summer school effects for lower income early-elementary students (Borman & Dowling, 2006; Borman, Goetz, & Dowling, 2009; Schacter & Jo, 2005). In these studies, summer-school students tended to either outgain their control-group peers during the summer vacation period or produce higher scores at the beginning of the following academic year. However, reported results were somewhat inconsistent as the statistical significance and size of the summer school effect varied across outcome measures, times of measurement, and analytic models applied to the data.

The literatures surrounding summer school and summer learning provide educational leaders with a logical and a developing empirical basis for offering lower performing or less-advantaged students a supplemental-support program (Alexander et al., 2001; Cooper et al., 1996, 2000; McCombs et al., 2011). For administrators and teachers, the targeted delivery of summer instruction is sensible from both a pedagogical and resource allocation perspective. However, by virtue of using a need-based delivery model, evaluative comparisons between students who participate in school-based summer instruction and those who do not can be misleading as the summer performance of students who choose to attend summer school may be stronger than similarly achieving peers who are invited to summer school but decline to attend or fully participate (Borman, Benson, & Overman, 2005). Conversely, the performance of summer school attendees may be lower than their peers whose academic year achievement status was strong enough to prevent a summer school invitation from being offered. In either case, the non-random selection of students into conditions makes the unbiased estimation of summer program effects difficult (Cooper et al., 2000).

In light of the evaluative challenge presented by non-equivalent control group designs, the current study was designed to implement a randomized field experiment in order to ascertain the efficacy of a summer literacy intervention delivered in one Pacific Northwest school district. A randomized design was implemented at the study site as prior investigation of the literacy intervention using quasi-experimental regression discontinuity methods indicated that program participation not only offset summer reading losses but was associated with a summer reading gain (Zvoch & Stevens, 2011). More specifically, at the cutscore for program admission, summer oral reading fluency gains were positive and approximately .40 of a standard deviation higher for students who attended summer school relative to their non-participant peers. However, despite the use of regression discontinuity models to account for the need-based assignment of students, concerns related to the "fuzziness" of the assignment cutscore as well as the generality of inferences across a wider reading proficiency range provided impetus for further investigation. In the following study, a series of inferential models was used to estimate the change in summer literacy performance associated with the summer assignment and participation status of a sample of kindergarten and first grade alumni. The study was designed to examine whether: (a) children assigned to summer school acquired literacy skills at a faster rate than comparable children not assigned to summer school, (b) participation in summer school was requisite to promote differential summer literacy gains, and (c) the application of different analytic models resulted in different summer learning estimates.

1. Method

1.1. Data source

Data analyzed in the present study were obtained in conjunction with a university-school district collaboration designed to foster a thorough formative and summative evaluation of a district-sponsored summer literacy intervention. Both the university and the school district are located in a moderately sized city in the Pacific Northwest. The school district serves close to 6000 students each year. The student body is approximately 75% White, 14% Latino, 3% African American, 3% Asian American, 3% Native American, and 2% Other. In recent years, approximately 44% of district students have received a free- or reduced-price lunch and 3% of district students have been identified as English language learners.

1.2. Summer program

Each summer, the district offers a five-week academically intensive literacy program to early-elementary students during the middle of the three-month summer vacation period. The program is designed to close the performance gap between strong and struggling readers and to ensure that struggling readers gain the foundational skills requisite to meet reading proficiency targets. The district focuses the summer program on the development of early literacy skills as proactive intervention efforts in beginning reading are viewed as the primary way to address and prevent initial reading difficulties from progressing into long-term reading and scholastic failure (Fuchs & Fuchs, 2006; Gersten et al., 2008; Torgesen, 2002). For school personnel, the focus and intensity with which best practice reading instruction can be delivered in a summer school setting is a distinct advantage. Unlike the academic year where a range of subjects are covered in shorter daily periods, the summer school setting provides educators with a unique opportunity to provide lengthy and exclusive instruction in the core literacy components that are fundamental to the attainment of successful reading outcomes.

The supplemental literacy program runs over a five-week period during the middle of the three-month summer vacation period. Instruction occurs 3.5 h/day, four mornings per week in small-class size environments housed in one central school site. The instruction is designed to closely align with the fundamental "big ideas" and best practices that underlie and facilitate early childhood literacy development (National Reading Panel, 2000). Students receive a minimum of 2 h of teacher-directed daily literacy 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). Daily lessons begin in a whole-group setting where attendance is taken, homework is checked, and a morning seat work packet is delivered. Literacy skills are then modeled and practiced primarily in small group (3–5 students) settings. Students are grouped based on skill level to allow small-group instruction to be maximally supportive and aligned with individual student needs. In each whole- and small-group arrangement, teachers provide direct modeling of literacy skills and offer multiple opportunities for practice with corrective feedback.

1.3. Analytic samples

Program effects were estimated from literacy data obtained from two student samples. The samples were comprised of students that either finished kindergarten ($N = 46$) or first grade ($N = 47$) during the 2009–2010 school year and completed district literacy assessments during the spring and fall of 2010. The two samples contained relatively equal percentages of students from special populations. The percentage of boys (55%), English language learners (15%) and free-lunch recipients (63%; a SES proxy) were equivalent while the percentage of students from ethnic-minority groups ranged from 19% in the kindergarten sample to 25% in the first-grade sample.

The student samples were generally more disadvantaged than the larger student body as a result of the selection of students identified as at moderate-risk for developing future reading difficulties. However, by virtue of the manner in which students were identified for assignment (see below), the district's most-challenged students were not included in either the kindergarten or first-grade sample. As a result, the struggling early-elementary school readers studied herein were not as low performing as those examined in other studies that have focused exclusively on students at greatest risk of negative summer outcomes (e.g., Borman & Dowling, 2006). It should also be noted that despite the relatively high percentage of free-lunch recipients included in the samples, the poverty conditions found in the community where the study was conducted are not as extreme as those encountered in larger urban environments. The implication is that students in the current samples have an academic and economic-risk profile that is somewhat more moderate than typically found in other need-based summer school investigations. Details regarding the criteria used to select and assign students are presented below.

1.4. Outcome measures

As a means for identifying, monitoring, and effectively intervening with students at heightened risk for poor literacy outcomes, the district administers a series of grade-specific assessments each academic year. In kindergarten and first grade, the district administers select subtests of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) and/or the Test of Oral Reading Fluency (TORF; Children's Educational Services, 1987). The DIBELS nonsense word fluency (NWF) subtest is administered to kindergarten students in January and May. In first grade, the NWF is administered three times (September, January, May). The TORF is added to the assessment schedule of all first-grade students during the January assessment window. Thereafter, the TORF is administered in September, January, and May each academic year.

Both DIBELS' NWF and the TORF are standardized, individually administered and age-appropriate early literacy skill assessments that have been developed to identify struggling readers and allow the monitoring of student progress over time. DIBELS' NWF is a one-minute assessment designed to measure students' understanding of the alphabetic principle. Students are presented with randomly ordered nonsense words (e.g., vad, ab) and are asked to orally reproduce the letter sound combinations. The final score is the number of correct letter sounds produced in one minute. The test developer reports a one-month alternate-form reliability estimate of .83 in one student sample and a multi-probe reliability estimate of .98 in a second (Good et al., 2004). Concurrent and predictive relationships with a variety of early reading measures have also been reported. In a sample of scores obtained from kindergarten children from a large urban school district, kindergarten NWF scores concurrently correlated .36–.56 with scores on subtests of the Test of Early Reading Ability and later correlated .50, .57, and .55 with scores obtained at the end of first grade on the TerraNova reading, vocabulary, and language subtests, respectively (Rouse & Fantuzzo, 2006). The test developer also reported that winter of first-grade NWF scores correlated .66 with spring of first grade scores on the Woodcock-Johnson Psycho-Educational Battery Readiness Cluster (Good et al., 2004).

The TORF is designed to measure students' accuracy and fluency with connected text. Student performance is measured by having students read each of three passages aloud for one minute. Word omissions or substitutions and hesitations of more than three seconds are scored as errors. If a word is self-corrected within a three-second timeframe, the response is scored as accurate. The median correct words-per-minute from the three passages serves as 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 while alternate-form 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 (Good & Jefferson, 1998) while 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 et al., 2004).

Student NWF and TORF performance were used both to determine summer school placement and to evaluate the efficacy of summer instruction. In the models presented below, the change in literacy performance between the final assessment in one academic year (May) and the first assessment in the next (September) served as the outcome. For the kindergarten sample, evaluations were conducted using the difference between spring of kindergarten and fall of first grade NWF scores while the difference between spring of the first grade and fall-of-second-grade TORF scores served as the outcome for the first-grade sample.

1.5. *Assignment criteria and participation rates*

Historically, the district has offered a need-based summer school placement to students who fall below benchmark performance on the spring literacy assessment or, in some specialized cases, are invited through teacher recommendation and district literacy coordinator approval. During kindergarten, students who produce fewer than 25 nonsense words-per-minute on the spring NWF subtest are invited to participate in the summer literacy program. For first grade students, an invitation is extended to those who produce fewer than 30 correct words-per-minute on the spring administration of the TORF. The kindergarten and first grade benchmarks have been established for conceptual and practical reasons. Conceptually, below benchmark students are considered to be at heightened risk for developing future reading difficulties (see Good & Kaminski, 2002). For example, first-grade students producing fewer than 30 words-per-minute are >20 words-per-minute below the norm the district uses to define an established reader. The consequence of achieving established reader status is realized in part when students later take the state reading proficiency test in third grade as 95% of established readers obtain a passing mark. Budgetary constraints have also allowed summer placements to be offered only to students most in need of supplementary instruction.

In 2010, the summer program was systematically expanded through a collaborative university–school district partnership in order to examine the effect of offering supplemental instruction to a wider range of students. At each grade level, an interval with a lower bound at the grade-specific spring benchmark (kindergarten NWF = 25, first grade TORF = 30) historically used by the district was first identified. Then an upper bound was established on the basis of student risk levels and the funding available for program expansion. Given available resources, a target of 25 additional students per grade was established and corresponding intervals that contained approximately 50 students were identified. For kindergarten, the interval ranged from a lower bound of 25 to an upper bound of 35 nonsense words-per-minute. For first grade, the interval ranged from a low of 30 to a high of 50 words correctly read-per-minute. Students projected to score within each of these intervals on their respective spring assessment formed the assignment pool. Within the pool, students were randomly assigned to receive an invitation to summer school or to serve in a no-treatment control group. It should also be noted that all students who scored below the historic kindergarten and first-grade cutpoint for program admission were invited to participate in summer school as well. Cutscore eligible participants were instructed in separate groups.

Across the kindergarten and first-grade samples, 64% ($N = 30$) of students in the assignment pool who received an invitation to summer school chose to participate. The participation rate was lower among kindergarten students ($n = 13$, 54%) than first-grade students ($n = 17$, 74%). Table 1 presents the pre- and post-test means and standard deviations by sample and summer assignment and participation status. Mean comparisons between the student groups (i.e., participants, refusers, controls; see below) on the preprogram spring literacy assessment revealed no statistically significant group differences for the kindergarten, $F(2, 43) = 1.61$, $MSR = 74.23$, $p > .05$ or first-grade student samples, $F(2, 44) = 0.29$, $MSR = 40.84$, $p > .05$.

Table 1
Pre- and post-test and summer change means and standard deviations by sample and summer treatment status ($N = 93$).

Variable	Participants		Refusers		Controls	
	M	*SD*	*M*	*SD*	*M*	*SD*
Kindergarten sample						
Spring NWF kindergarten	29.77	9.99	34.55	8.31	34.95	7.88
Fall NWF grade 1	41.77	15.12	29.45	7.75	30.27	11.58
Summer change	12.00	16.03	−5.09	10.73	−4.68	12.01
N	13		11		22	
First grade sample						
Spring TORF grade 1	48.76	12.65	47.00	9.59	50.71	11.97
Fall TORF grade 2	56.24	13.95	43.17	9.62	46.04	15.34
Summer change	7.48	8.25	−3.83	6.94	−4.67	13.07
N	17		6		24	

Note. Participants and refusers were students who were randomly assigned to summer school. Students who were not randomly assigned to summer school were controls.

1.6. *Predictor variables*

For each grade, dummy codes were used to form two summer status indicators to estimate and compare student literacy outcomes. The first indicator distinguished between students who were randomly assigned to summer school (whether they attended or not) and their peers who were in the assignment pool, but did not receive an invitation to attend (i.e., the control group). This contrast provided an Intent-to-Treat (ITT) estimate of the summer school effect at each grade level. A second set of dummy codes was then used to distinguish between treatment participants, refusers, and students in the control group. Two dummy codes were used to represent the three groups with the control group serving as the referent or comparison group. This specification facilitated the estimation of a Treatment-On-Treated (TOT) contrast between

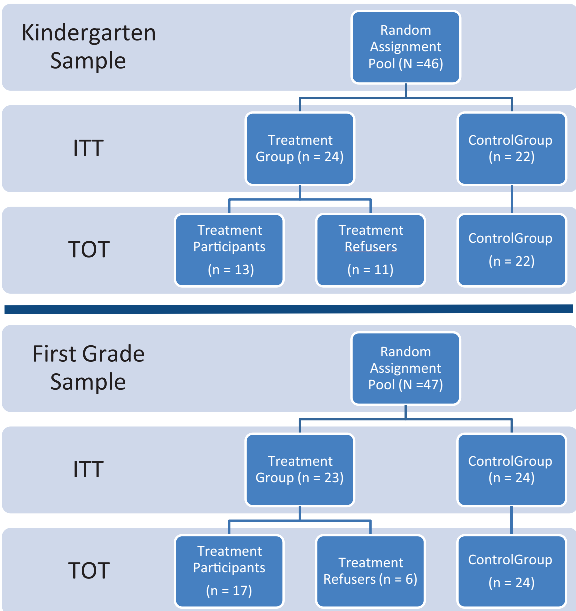

Fig. 1. Treatment status classification, group size, and analytic contrasts for the kindergarten and first grade samples.

treatment participants and control group students. Fig. 1 presents a flowchart that represents each treatment status classification, associated group size, and analytic contrasts for each student sample.

1.7. Analytic procedures

A series of ordinary least squares (OLS) regression models was estimated to identify the extent to which assignment to and participation in summer school was associated with students' summer learning. Separate models were run at each grade level as the summer performance of students was evaluated using different literacy outcome measures (i.e., DIBELS' NWF at kindergarten, TORF at first grade). OLS regression models were estimated rather than multilevel models as students in the control group were not nested in an identifiable higher unit of organization during the summer. The first regression model yielded a conservative ITT estimate contrasting the summer gains of those assigned to treatment (whether they participated or not) with the summer gains of the control group. The ITT analysis was used to provide a policy-relevant estimate of the treatment effect that takes into account that not all of the students will participate when a given program is offered. The second regression model (i.e., TOT) yielded group specific contrasts between treatment participants and controls and between treatment refusers and controls. In the reported analyses, group difference estimates are provided in raw and standardized (Hedges' *g*) units.

After estimation of the OLS models, a supplemental analysis was conducted to further evaluate the effects of summer school participation and to correct for potential bias in the estimation of program effects resulting from treatment non-compliance (Angrist, Imbens, & Rubin, 1996; Sheiner & Rubin, 1995; West & Sagarin, 2000). In the supplemental analysis, weighting methods were used to correct the ITT estimates derived from the OLS regression models. Weighting the ITT estimates by the proportion of treatment group members who actually received treatment (i.e., compliance-adjusted ITT = $(M_{\text{assigned}} - M_{\text{control}})/p_c$, where p_c = the proportion of compliers), the impact of "no shows" (students who were assigned to treatment but refused to participate in summer school) was taken into account (Bloom, 1984; Gennetian, Morris, Bos, & Bloom, 2005).

Table 2
Intent-to-treat (ITT) and treatment-on-treated (TOT) kindergarten regression model results (N = 46).

	ITT		TOT	
	b	*ES*	*b*	*ES*
Intercept	−4.68		−4.68	
	(3.05)		(2.71)	
Assigned to Tx	8.85*	0.60		
	(4.22)			
Participant			16.68*	1.17
			(4.55)	
Refuser			−0.41	−0.03
			(4.80)	
R^2	.09		.27	

Note. The referent group in the ITT and TOT models consisted of students who were randomly assigned to the control group. Standard errors are in parentheses; *ES* = effect size.
* $p < .05$.

2. Results

2.1. Kindergarten models

Results associated with the kindergarten models are presented in Table 2. In the ITT model contrasting students that were assigned to treatment with students assigned to the control group, the average change in summer nonsense word fluency for control students was a statistically non-significant 4.68 nonsense word loss ($p > .05$). For those assigned to summer school regardless of actual participation, the average change in summer nonsense word fluency was estimated as a 4.17 word-per-minute gain (i.e., 8.85–4.68 = 4.17). The coefficient associated with the between-group contrast was statistically significant ($b_1 = 8.85$, $p < .05$). Computation of Hedges' *g* showed that the 8.85 word-per-minute group difference represented an intent-to-treat effect of .60 of a standard deviation (i.e., $g = b_1/SD_p = 8.85/14.83 = .60$).

In the TOT model, which compared all three groups of students, the contrast between nonparticipating students (i.e., refusers versus controls) was not statistically significant ($p = .932$). However, the contrast between treatment participants and students randomly assigned to the control group was statistically significant with a between-group treatment difference of 16.7 nonsense words-per-minute and an associated effect size of 1.17 standard deviations. Further, when the regression was rerun with the group of treatment participants serving as the referent, the summer mean gain in NWF for participants ($b_0 = 12.00$) was statistically different from zero ($p < .05$). Overall, the ITT and TOT models respectively accounted for 9% and 27% of the variation in the change in summer NWF performance.

In models designed to evaluate possible relationships between student characteristic variables and the change in NWF scores, none of the demographic indicators (i.e., gender, ethnicity, free lunch or English language status) were a statistically significant predictor of the change in nonsense word fluency. Spring NWF scores were negatively related (ITT model = −0.81; TOT model = −0.65) to the change in summer NWF ($p < .05$). However, with the exception of the assignment variable in the ITT model ($b_1 = 7.41$, $p = .084$), the size and direction of the summer treatment coefficients and the associated patterns of statistical significance in the covariate-adjusted models were otherwise the same as those presented in Table 2.

2.2. First-grade models

Results associated with the first-grade models are presented in Table 3. In the ITT model contrasting the treatment and

Table 3
Intent-to-treat (ITT) and treatment-on-treated (TOT) first grade regression model results ($N = 47$).

	ITT		TOT	
	b	*ES*	*b*	*ES*
Intercept	−4.67 (2.32)		−4.67* (2.23)	
Assigned to Tx	9.19* (3.32)	0.78		
Participant			12.14* (3.47)	1.03
			0.83 (4.99)	0.07
R^2	.15		.23	

Note. The referent group in the ITT and TOT models consisted of students who were randomly assigned to the control group). Standard errors are in parentheses; *ES* = effect size.
* $p < .05$.

control groups, the average change in summer oral reading fluency for control students was a statistically non-significant 4.67 word-per-minute loss in fluency ($p = .051$). For those randomly assigned to summer school, the average change in summer oral reading fluency was estimated as a 4.5 word-per-minute gain (i.e., 9.19–4.67 = 4.52). The coefficient associated with the between-group contrast was statistically significant ($b_1 = 9.19$, $p < .05$). Computation of Hedges' *g* showed that the 9.19 word-per-minute group difference represented an intent-to-treat effect of .78 of a standard deviation (i.e., $g = b_1/SD_p = 9.19/11.79 = .78$).

In the TOT model, the contrast between treatment refusers and control students was not statistically significant ($p = .867$). However, contrasting treatment participants with students randomly assigned to the control group revealed a statistically significant between group treatment difference of 12.14 words-per-minute over the summer with an associated effect size of 1.03 standard deviations. When the regression was rerun with the group of treatment participants serving as the referent group, the summer TORF gain for participants ($b_0 = 7.47$) was statistically different from zero ($p < .05$). Fig. 2 presents the change in oral reading fluency between the end of first grade and the beginning of second grade for each of the three student groups. While all groups had statistically equivalent mean ORF scores at the end of first grade, the trajectory of oral reading fluency gains was positive for summer school participants, but negative for both refusers and control group students. Overall, the ITT and TOT models respectively accounted for 15% and 23% of the variation in the change in summer ORF performance.

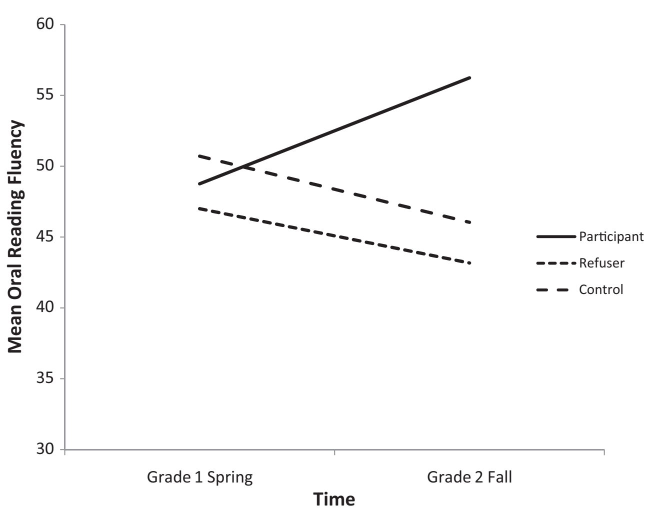

Fig. 2. Mean oral reading fluency as a function of summer treatment status and time.

In models that contained student characteristic variables, student ethnicity in the ITT analysis was the only demographic indicator that was a statistically significant predictor of the change in summer oral reading fluency. Relative to their counterparts, ethnic minority students were predicted to lose approximately 9 words per minute over the summer ($p < .05$). However, with the exception of the assignment variable in the ITT model ($b_1 = 6.03$, $p = .099$), the size and direction of the summer treatment coefficients and the associated patterns of statistical significance in the covariate-adjusted models were otherwise the same as those presented in Table 3. It should also be noted that one influential case was identified and excluded from the first grade analyses. The case in question was a student who was assigned to summer school but did not attend. This student recorded the largest gain in oral reading fluency over the summer in the entire sample (25 words per minute). The standardized DfBeta associated with this case (.88) was discontinuous from the rest of the distribution by over a half of a standard deviation and greatly exceeded the influence threshold of $3/\sqrt{47} = .44$ (Stevens, 2009). The associated unstandardized DfBeta indicated a positive change in slope of 4.12 words per minute for the group of refusers if the case were to be retained. Inclusion of this case did not change the pattern of statistical significance in any of the regression models, but did inflate and misrepresent the normative gain of summer school refusers.

2.3. Supplemental analysis

To further evaluate the effects of summer school participation and to correct for potential bias in the estimation of program effects resulting from non-compliance with treatment assignment (Angrist et al., 1996; Sheiner & Rubin, 1995; West & Sagarin, 2000), correction methods were applied to the ITT estimates presented in Tables 2 and 3 to derive compliance-adjusted estimates of the summer school effect (Bloom, 1984; Gennetian et al., 2005). Weighting the ITT estimates by the proportion of treatment group members who actually received treatment, the estimated program effect for the kindergarten sample was a statistically significant 16.34 additional nonsense words-read-per-minute for summer school participation, $t(44) = 2.32$, $p = .013$, Hedges' $g = 0.75$. Adjusting for non-compliance in the first grade sample, the estimated program effect was a statistically significant 12.45 additional words-read-per-minute for summer school participants, $t(45) = 2.93$, $p = .003$, Hedges' $g = 0.88$.

3. Discussion

As a means for arresting the slide in achievement growth that is commonly observed among at-risk students during the summer vacation period, many school districts have implemented summer instructional programs (McCombs et al., 2011). The targeted delivery of summer school is often viewed as a cost-effective approach for providing an additional dose of instruction during a period of time when hard-fought academic year gains can be lost. However, by virtue of the non-random manner in which students are selected to receive a summer school placement, unbiased summative evaluation of summer school effects is often challenging (Cooper et al., 2000). In the current study, a unique opportunity for evaluating the efficacy of a five-week summer literacy intervention arose through a collaborative partnership formed between university researchers and school district personnel. Employing a field-based randomized trial, the effect of assignment to and participation in summer school was estimated in a series of inferential statistical models.

Intent-to-treat (ITT) analyses were first conducted to compare all students who were randomly assigned to summer school (whether or not they participated) to their peers. Results indicated that the group of students assigned to summer school outperformed their peers on DIBELS' NWF by approximately .60 of a standard deviation between kindergarten and grade 1 and .78 of a standard deviation on the TORF between grades 1 and 2. Differences in summer performance were more pronounced when actual participants (i.e., those who were assigned and participated in summer school) were compared with students in the control group. In the TOT analyses, the between-group difference in summer literacy gains was approximately 1 standard deviation in the first grade sample and 1.17 standard deviations in the kindergarten sample. Supplemental compliance-adjusted analyses that corrected for the potential bias associated with non-compliance among the group of students assigned to summer school yielded treatment effect estimates intermediate between the ITT and TOT estimates.

Overall, students in the control group showed mean declines of approximately 5 words-per-minute over the summer in both samples (NWF scores in the kindergarten sample, TORF scores in the first-grade sample), while participants had fluency gains of approximately 12 (nonsense) words-per-minute in the summer following kindergarten and 7.5 words-per-minute in summer following first grade. Of additional note, no learning differences were observed between students who refused a summer placement and students assigned to the control group, suggesting that assignment to summer school in and of itself (i.e., intent-to-treat) was not sufficient to promote summer gains. From a practical standpoint, the summer gains of participants were large enough to place these students firmly in the low risk (NWF > 25) or established reader (TORF > 50) proficiency category at the beginning of the next academic year. These findings thus support the experiential intuition of school district personnel regarding the benefits of summer school and generally suggest that targeted summer instruction can be an effective strategy to support student learning over the summer months.

The performance gap that emerged in literacy scores at the start of the new academic year between moderately at-risk readers with different summer treatment classifications (i.e., participants, refusers, and controls) is an indicator of the potential that summer instruction holds for those who participate in a school-based supplemental support program. However, questions regarding the absolute and relative effects of providing treatment to more-advantaged recipients have been raised. In particular, concern has been expressed that the delivery of treatment to higher performing or less-disadvantaged groups of students may serve to further exacerbate preexisting performance gaps (Ceci & Papierno, 2005). In the current study, comparative analyses (not presented) conducted on data obtained from high-risk students who were invited to and participated in summer school as a result of scoring below the literacy cutscore for program admission (kindergarten NWF < 25; first grade TORF < 30) support this contention. In brief, the mean literacy gap between the moderate (i.e., randomly assigned summer school participants) and the highest risk summer school participants increased from 15 to 22 nonsense words-per-minute from the end of kindergarten to the beginning of first grade (moderately at-risk reader advantage) while the gap in oral reading fluency increased from 27 to 34 words-per-minute from the end of first grade to the beginning of second grade (moderately at-risk reader advantage). Findings that demonstrate an increase in the achievement gap between moderate and high-risk students suggests that educational leaders should be cautious in the allocation of summer placements if a central goal of the provision of supplemental instruction is to help lower achieving students keep pace or close the gap with their more advantaged peers.

Overall, the summer literacy gain experienced by summer school participants and the literacy loss experienced by their non-treated peers was generally consistent with meta-analytic findings documenting summer school effects and the summer learning outcomes of children and early adolescents (Cooper et al., 1996, 2000). The pattern of results in this study was also largely consistent with findings associated with two other recent randomized field investigations that have examined the effects of supplemental summer instruction in early-elementary school student samples (Borman et al., 2009; Schacter & Jo, 2005). Similar to the kindergarten and first-grade students who were randomly assigned to summer reading programs in Los Angeles (Schacter & Jo, 2005) and Baltimore (Borman et al., 2009), treatment participants in the current samples gained literacy skills over the summer and demonstrated higher proficiency on a key literacy measure (relative to their control-group peers) at the beginning of the following academic year. Moreover, as with the two summer reading programs implemented with economically-disadvantaged children in large urban environments (Borman et al., 2009; Schacter & Jo, 2005), demographic background characteristics were also generally not predictive of summer literacy outcomes among moderately at-risk readers in a medium-sized city in the Pacific Northwest.

Current results were also generally consistent with those reported in conjunction with a randomized field experiment conducted among another sample of low-income children in Baltimore (Borman & Dowling, 2006). In the Borman and Dowling (2006) study, kindergarten and first-grade students from 10 high-poverty schools were randomly assigned to the Teach Baltimore Summer Academy, a seven-week, academically rigorous summer program. Although a treatment assignment (i.e., ITT) effect was not obtained, an effect of treatment receipt was reported after statistical models that provided an adjustment for treatment noncompliance were applied to the intervention data. In particular, careful consideration of the participation status of students with Complier Average Causal Effect (CACE) models enabled the identification of a treatment effect on each reading outcome in the Borman and Dowling (2006) study while estimation of TOT and compliance-adjusted ITT models revealed larger estimated treatment effects in the current study. These findings suggest that assignment to summer school in and of itself is not sufficient to positively affect the learning outcomes of students who refuse treatment (also see Borman et al., 2005). Moreover, in samples with large noncompliance rates, treatment estimates that underestimate program impact are likely. As a result, statistical models that adjust for noncompliance within the intervention group are recommended as a strategy to obtain more accurate treatment effect estimates (Angrist et al., 1996; Imbens & Rubin, 1997; Jo, 2002; Little & Yau, 1998).

Although largely consistent with other recent investigations, the size and strength of the summer reading fluency outcomes reported herein should be tempered with respect to the nature of the instructional intervention delivered to students. As described earlier, the summer literacy intervention was designed to facilitate the delivery of an intense daily dose of direct literacy instruction to students. Students received explicit, fluency-focused instruction primarily in small group environments by highly trained and experienced instructors who utilized best practice teaching strategies. The frequent modeling of skills, regular provision of guided student practice opportunities, and delivery of real time individualized feedback in a scripted instructional framework that formed the core of this intervention program was derived from and is designed to be consistent with National Reading Panel (2000) recommendations. Moreover, with a highly trained and experienced instructional faculty delivering highly scripted literacy lessons, relatively little deviation from program protocol was observed either within or between-instructors during treatment integrity checks conducted over the five-week instructional period (Zvoch, in press).

The strength and consistency of the instructional dose delivered to students thus distinguishes the summer program examined here from others that historically have provided remedial instruction or instruction disconnected from the academic year curriculum or otherwise may have been less targeted and aligned with student needs (see Cooper et al., 2000).

The congruence between the fluency-development instructional emphasis and the fluency-based measures that served to index the summer learning gains and losses of students also distinguished the current study from other summer school investigations. More specifically, the exclusive use of assessments sensitive to changes in reading fluency and closely aligned with the instructional model may have contributed to the discrepancy in the size of reported effects. In other words, with the use of outcome measures (NWF & TORF) that are directly proximal to program activities, estimated effects were likely larger and less reflective of a children's total reading ability than those that would have followed had individual change on a somewhat more distal and comprehensive reading measure (e.g., a norm-referenced reading comprehension or vocabulary measure) been evaluated (Lipsey & Cordray, 2000). The close alignment between the fluency outcomes, a key but non-comprehensive measure of a child's reading ability, and the intervention's instructional focus is thus a distinct limiting feature of the current study.

Another distinct limiting feature of the current study is the size and nature of the analytic sample(s). With fewer than 100 moderately at-risk participants, the sample size was small and not typical of most studies of summer programs. For example, whereas the Teach Baltimore Summer Academy sample ($N = 686$; Borman & Dowling, 2006) was comprised entirely of economically disadvantaged inner-city students, the students included in the present investigation came from a mix of economic backgrounds and were not at the highest level of reading risk. Children from somewhat more-advantaged academic and economic backgrounds may have relatively more access to the home and community-based educational resources that facilitate out-of-school learning (Burkam et al., 2004; NCES, 2004). In cases where these resources can be combined with effective school-based summer instruction, the benefits associated with an academically focused summer school program may be enhanced further (Cooper et al., 2000). Replication of the current results with larger, more diverse samples on comprehensive and instructionally aligned reading measures would thus be reassuring.

Taken together, the implementation and evaluation of a school-based field trial with randomization accomplished at the student level is currently somewhat unique in the field of education despite the policy interest and funding priorities of the U.S. Department of Education (see Boruch & Mosteller, 2002; Burtless, 2002; Institute of Educational Sciences [IES], 2008). Where feasible, IES has encouraged and supported the use of the randomized trial as a means for providing direct control over the universe of potential confounding variables (IES, 2008). Although small in scale, results obtained from two samples of early-elementary children identified as at-moderate risk for future reading difficulty and randomly assigned to summer school suggest that literacy instruction delivered during the summer months can positively impact the reading fluency gains of struggling readers. The positive results associated with the probabilistic assignment of students to conditions serves to strengthen the inferences that can be drawn regarding the effects of providing a strong dose of summer instruction (also see Denton, Solari, Ciancio, Hecht, & Swank, 2010). However, the unique characteristics of the study context, intervention and measurement characteristics, and the student samples caution against drawing general conclusions regarding the potential strength of summer instruction delivered to students from different backgrounds (e.g., age, socio-economic status, achievement level) in other scholastic and community contexts. Instead, the size of the effects associated with the current study are most likely at the upper bound of what can reasonably be expected from the adoption and implementation of a rigorous school-based summer instructional intervention delivered to moderately at-risk early elementary students with great fidelity. Replication with larger, more-representative samples and inclusion of more-comprehensive reading measures would thus serve to better contextualize the findings reported herein. Additional study with diverse methods, measures, and samples is therefore urged to more clearly identify the conditions under which summer instruction can serve to promote summer learning gains.

The more frequent use of the randomized experiment and strong quasi-experimental designs like regression discontinuity (RD) and interrupted-time series approaches across different samples and diverse contexts will serve to strengthen inferences regarding the size and direction of the summer school effect. However, as long as participation in summer school is voluntary, researchers will need to pay particular attention to the number and characteristics of students who decline a summer placement. In the current study, summer school refusers were not systematically different than their peers who chose to participate, and refusers experienced summer outcomes similar to students who were not assigned to receive summer school instruction. In other populations, the pre-treatment equivalence of groups and the associated pattern of summer outcomes may differ substantially. It is thus recommended that various ITT, TOT, and compliance-adjusted estimates be presented in any report that is based on a randomized design with non-compliance or an RD design that does not maintain a sharp cutpoint assignment protocol (e.g., as a result of misallocation and/or non-compliance). In these instances, an additional data collection that captures the degree of access to and utilization of educational resources in the home and community would also serve to better contextualize the summer learning outcomes of treatment participants, refusers, and controls. With continued application of research designs that support unbiased estimation of summer treatment effects across a range of diverse intervention contexts, educational leaders will be better positioned to consider the costs and benefits associated with offering a local program of summer instructional support.

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.

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

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444–455.

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.

Bloom, H. S. (1984). Accounting for non-compliers in experimental evaluation designs. *Evaluation Review*, *8*, 225–246.

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.

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.

Boruch, R., & Mosteller, F. (2002). Overview and new directions. In F. Mosteller, & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 1–14). Washington, DC: Brookings Institution Press.

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.

Burtless, G. (2002). Randomized field trials for policy evaluation: Why not in education? In F. Mosteller, & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 179–197). Washington, DC: Brookings Institution Press.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have nots" gain but the "haves" gain even more. *American Psychologist*, *60*(2), 149–160.

Children's Educational Services. (1987). *Test of oral reading fluency (TORF)*. Eden Prairie, MN: Author.

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.

Denton, C. A., Solari, E. J., Ciancio, D. J., Hecht, S. A., & Swank, P. R. (2010). A pilot study of a kindergarten summer school reading program in high poverty urban schools. *Elementary School Journal*, *110*(4), 423–439.

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.

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.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*(1), 93–99.

Gennetian, L. A., Morris, P. A., Bos, J. M., & Bloom, H. S. (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 75–114). New York, NY: Russell Sage Foundation.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. (NCEE 2009-4045. 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, NY: Guilford Press.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Education Achievement.

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

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., & Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with non-compliance. *The Annals of Statistics*, *25*, 305–327.

Institute of Education Sciences. (2008). *What Works Clearinghouse: Evidence standards for reviewing studies, version 1.0*. Washington, DC: Author.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, *27*, 385–409.

Little, R. J., & Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods*, *3*, 147–159.

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology*, *51*, 345–375.

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

National Center for Education Statistics (NCES). (2004). *The summer after kindergarten: Children's activities and library use by household socioeconomic status*. U.S. Department of Education, Institute of Education Sciences. (NCES 2004-037).

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).

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

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/).

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the dynamic indicators for basic early literacy skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review*, *35*(3), 341–355.

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.

Sheiner, L. B., & Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapy*, *57*, 6–10.

Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York, NY: Routledge Academic.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement (Research Rep. 109)*. Minneapolis, MN: 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.

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

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.

West, S. G., & Sagarin, B. J. (2000). Participant selection and loss in randomized experiments. In L. Bickman (Ed.), *Research design: Donald Campbell's legacy* (pp. 117–154). Thousand Oaks, CA: Sage.

Zvoch, K. (2009). A longitudinal examination of the academic year and summer learning rates of full and half-day kindergartners. *Journal of Education for Students Placed At Risk*, *14*(4), 311–333.

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. (in press). How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. American Journal of Evaluation.