Studies in Educational Evaluation 55 (2017) 145–152

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Multivariate summer school effects

Keith Zvoch[a,*], Mark C. Robertson[b]

[a] Department of Educational Methodology, Policy and Leadership, University of Oregon, United States
[b] Concordia University, Portland, College of Education, Centennial Hall 012, 2811 NE Holman Street, Portland, OR 97211, United States

ARTICLE INFO

Keywords:
Summer school
Summer learning
Randomized field trial

ABSTRACT

The effect of assignment to and participation in summer school for a moderately at-risk sample of kindergarten students was examined with multivariate analytic methods. A multivariate analysis of variance applied to difference scores capturing the change in summer literacy outcomes revealed that kindergarten students randomly assigned to summer school outperformed their control group peers on a linear composite of early literacy indicators. The estimated group difference was greater when participation in summer school was distinguished from receipt of the summer program offer in analyses that explicitly adjusted for the proportion of students who failed to comply with their assignment. These results demonstrate that the nature and generalizability of the inference regarding program performance varies in relation to the intended and achieved design and the analytic model applied to data. Implications for the evaluation of summer school programs are discussed.

1. Introduction

Summer recess poses a challenge for students struggling to acquire academic knowledge and skills (Atteberry & McEachin, 2016; Burkam et al., 2004; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). Over the summer months, the learning rate of economically advantaged students slows relative to the academic year while the performance of economically disadvantaged students and their peers at the lower end of the achievement distribution tends to stagnate or decline (Alexander, Entwisle, & Olson, 2001; Downey, von Hippel, & Broh, 2004; Downey, von Hippel, & Hughes, 2008; Heyns, 1987, 1978). The summer learning slowdown has been attributed to the reduction in formal and informal opportunities to engage with instruction and other learning activities that are conducive to building academic skills and academic patterns of thinking (Alexander et al., 2001; Cooper et al., 1996; Entwisle, Alexander, & Olson, 1997). Access to key educational resources (e.g., books, educational media, trips to the library) can be particularly limited for students from disadvantaged backgrounds as their families may not possess the financial means or hold and convey the academic expectations requisite to facilitate a culture of summer learning (Benson & Borman, 2010; Gershenson, 2013; Pallas, 2016; Slates, Alexander, Entwisle, & Olson, 2012). Awareness of the basis for and consequences of the “summer slide” has led many school districts to enact policies or programs that seek to provide an instructional scaffold during the summer months (Kim & Quinn, 2013; Kim & White, 2008; McCombs et al., 2014; Patall, Cooper, & Batts Allen, 2010). The purpose of this paper is to investigate the impact of one academically rigorous, school-based summer literacy program. In the following, we estimate and evaluate the effect of a targeted summer school program on the change in student performance on a weighted composite of early literacy skill indices.

Differential summer learning rates serve to further exacerbate preexisting achievement gaps between those less and more advantaged (Benson & Borman, 2010). Analysis of national and local survey data reveals that disadvantaged students continue to lose ground relative to their more advantaged peers despite similar academic year growth rates (Alexander et al., 2001; Downey et al., 2004, 2008; Heyns, 1987, 1978). The in- and out-of-school achievement growth pattern established in these studies demonstrates the benefit of school-based instruction, but also illustrates the structural challenges that schools face (Rambo-Hernandez & McCoach, 2015; von Hippel, 2009). Despite the efforts of school staff, disadvantaged students find themselves falling further behind their peers and increasingly unsuccessful in challenging subject matter. Compounded over time, students from disadvantaged backgroundsare less likely to take the advanced courses requisite for college preparation, and instead are more likely to disengage from and drop out of formal schooling (Alexander, Entwisle, & Olson, 2007). The summer disparity in learning has thus become a focal point for educational leaders seeking to keep their most at-risk students (i.e., low performing and/or economically disadvantaged students) on track to achieve positive short and longer term educational goals (McCombs et al., 2011).

To address the challenges that stem from a 3-month summer break

* Corresponding author at: University of Oregon, Department of Educational Methodology, Policy and Leadership, College of Education, 5267 University of Oregon, Eugene, OR 97403-5267, United States.
E-mail address: kzvoch@uoregon.edu (K. Zvoch).

http://dx.doi.org/10.1016/j.stueduc.2017.10.003
Received 4 May 2017; Received in revised form 8 October 2017; Accepted 9 October 2017
0191-491X/

from schooling, some school districts have chosen to adopt a year round schedule (NAYRE, 2007; Skinner, 2014) while others have implemented a targeted summer enrichment program for at-risk students (McCombs et al., 2011, 2014). Year round schools rearrange the traditional 180 day school calendar to provide for shorter breaks (e.g. 15–20 days of intersession for every 45–60 instructional days) in an attempt to keep the rhythm of instruction more constant (Cooper, Valentine, Charlton, & Melson, 2003). Notably, a rearrangement of the school calendar to a year round schedule affects factors proximal to students, teachers, and parents (e.g., time for teacher planning and professional development, complications with child care for parents), but does not increase the total amount of time available for instruction (Cooper, 2004; von Hippel, 2007, 2016). When academic outcomes in year round and nine month calendar schedules are compared, findings indicate that students in year round schools learn relatively faster during summer months (when in session) whereas students in nine month schools learn faster during the rest of the year, resulting in equivalent annual achievement gains (von Hippel, 2016).

In contrast, targeted summer programs serve as an extension to the school calendar and offer students who are most at-risk strategic supplemental instruction. Summer instructional programs or "summer school" are strategic in that educational resources can be used to target those most in need of support. Summer programs provide additional opportunity for students to receive content area instruction and practice academic tasks. As a result, students are more likely to increase their academic achievement in the content areas in which supplemental instruction is delivered (McCombs et al., 2011). More specifically, Cooper et al. concluded that economically disadvantaged students who participated in summer school were able to maintain or slightly improve upon academic gains from the prior year in reading and mathematics (Cooper, Charlton, Valentine, & Muhlenbruck, 2000). Other findings revealed a U-shaped relationship across grade levels as early-elementary and late-secondary students were the largest benefactors from program participation. The size of the summer school effect across all samples examined was approximately one quarter of a standard deviation. However, when disaggregated by economic background, the average effect for middle class students (d = 0.44) was larger than that observed from low income students (d = 0.20).

Although the literature on summer school indicates that students who attend summer programs generally benefit from the experience (Cooper et al., 2000; Kim & Quinn, 2013; but also see Augustine et al., 2016), many authors advise caution when interpreting summer program outcomes. For example, if a program's effectiveness is evaluated by pretest-posttest gain scores, then statistical regression to the mean may inflate observed gains and lead to an overestimation of program effects. Also, if a comparison group is not available, then it is not possible to measure the effect the program had on students relative to students not attending the program (Cooper et al., 2000; Stein & Fonseca, 2016). To begin to evaluate the efficacy of summer school, it is typically requisite that a control condition or a comparison group be available to contrast achievement outcomes. However, it should be noted that a summer program that offers placements on a need basis may also be serving students with a home environment that is concomitantly less conducive to academic success. If at-risk students are later compared to a group of students that did not receive an invitation (due higher performance), achievement outcomes are confounded by initial group differences and extraneous factors (e.g., home-based instruction and educational resource access) that are non-randomly distributed across households (Borman, Benson, & Overman, 2005; Stein & Fonseca, 2016). The difficulty in separating instructional effects from the distinct background characteristics of summer school participants thus often weakens the inferences drawn regarding summer program outcomes. As a result, researchers need to carefully consider the manner in which students are invited to participate and whether invitees choose to attend (Augustine et al., 2016; Borman & Dowling, 2006).

To mitigate the presence of extraneous factors that can distort treatment and control group comparisons, strong methodological approaches such as experimental designs with random assignment to conditions are recommended (Shadish, Cook, & Campbell, 2002). In situations where randomization is not feasible, strong quasi-experimental alternatives like the regression discontinuity or interrupted time series designs are encouraged (Murnane & Willett, 2011; Shadish et al., 2002). The use of designs where the selection mechanism is known and/or can be completely modeled enables the causal impact of summer school to be identified. As attendance at a summer program is often voluntary, researchers who study these programs using random assignment or other techniques should also be prepared to examine outcomes as a function of the compliance status of students. For example, in a study where over 600 early elementary students were randomly assigned to a 7-week summer school program, the experimental effect associated with treatment assignment was not statistically significant. However, in supplemental analyses that adjusted for noncompliance with treatment offer, a positive summer school effect for the complier subgroup was observed (Borman & Dowling, 2006).

In light of the evaluative challenge presented by nonrandom sorting into conditions, the current study utilized a randomized field experiment as a basis for ascertaining the efficacy of a summer literacy intervention. The intervention context was that of a 5-week summer program delivered to struggling early readers who completed kindergarten in the prior academic year. To implement the design, a screening assessment administered in the spring of the kindergarten year was used to identify a pool of students at moderate risk of future reading difficulty. These students were then randomly assigned to receive a summer school placement offer. As the summer program was focused on developing multiple beginning literacy components and skills (e.g., phonemic awareness, alphabetic understanding, fluency/automaticity), we used a multivariate analysis that permitted the simultaneous examination of the change in several literacy outcomes. Statistical adjustments were also made to account for noncompliance with treatment assignment in order to estimate the impact of summer school for those who attended. Overall, the study was designed to examine (a) whether students assigned to summer school outperformed comparable students not assigned to summer school, and (b) how the size and nature of estimated summer school effect changed when noncompliance with the treatment offer was taken into account.

2. Method

2.1. Data source

The study was conducted in a moderately-sized school district in the United States' Pacific Northwest. In a typical year, the district serves approximately 6000 students. During the implementation period, the student population was 74% White, 14% Latino, 3% African American, 3% Asian American, 3% Native American, and 2% other. Approximately 50% of students were eligible for a free or reduced priced lunch and 3% of district students were identified as English language learners.

2.2. Summer literacy program

Summer school was offered to kindergarten alumni for 5-weeks during the middle of the 3-month summer vacation period (mid-July to mid-August). Summer school was provided 3.5 h/day 4-days per week (Monday through Thursday) at a central school site. The summer program focused 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; National Early Literacy Panel, 2008; National Reading Panel, 2000; Torgesen, 2002). Each day, students received a minimum of two hours of teacher-directed daily literacy instruction in the critical beginning reading skills of phonemic awareness (oral blending and segmentation) and alphabetic understanding (letter-sounds, decoding, phonic analysis). Teacher modeling of skills and student opportunity to practice with formative feedback was an integral part of the instructional intervention.

Daily lessons began with students meeting as a whole classroom group for the purpose of taking attendance, reviewing homework, and engaging in a warm-up activity. Students then worked in homogeneous skill-based small groups of 3–5 so that instruction and support could be targeted to better meet each student's individualized needs. Upon completion of the first small group work session, students were given a short recess, followed by additional small group and individual work opportunities within literacy stations. Stations consisting of curriculum-identified tasks which focus on specific reading skills allowed teachers greater ability to address individual student needs through differentiated instruction. At the end of each daily session, students reconvened as a whole classroom group to review concepts addressed during the day's lesson.

Summer literacy instruction was provided by regular classroom teachers and highly trained educational assistants. Program administrators selected summer literacy instructors through a competitive skill-based application process. Instructors in the summer literacy program underwent additional professional development training prior to the start of the program. Small group instruction was monitored by trained observers during the second and fourth week of the 5-week intervention period. An observational protocol aligned with the fundamental "big ideas" and best practices that underlie and facilitate early childhood literacy development (National Reading Panel, 2000) was used during each observation period. Core components that were monitored were those common to scripted reading models, including the organization and sequenced use of instructional materials, the focus on basic phonemic, alphabetic, and fluency skills, correct lesson wording and pacing, appropriate modeling of decoding skills and strategies, allowance for student practice, and the provision of individual feedback to correct errors and positively reinforce correct student responses (Zvoch, 2012).

2.3. Analytic sample

Evaluation of the efficacy of the summer literacy program was made possible by implementation of a randomized control trial. The trial was conducted on a sample of students identified at moderate risk of future reading difficulty (see below). Summer school invitations were randomly extended to 24 of the 46 students included in the randomization pool (i.e., 24 students were assigned to the treatment condition, 22 students were assigned to the control condition).[1] Overall, the random assignment pool was 74% non-Latino White ($n = 34$) and 54% male ($n = 25$). Sixty-three percent of the sample received a free or reduced price lunch ($n = 29$) and 15% were identified as English language learners ($n = 7$). Summer school participation was voluntary, and not all students who received an invitation to summer school accepted the offer. Eleven of 24 students (46%) assigned to summer school chose not to attend. However, none of the students assigned to the control condition crossed over into the treatment condition. Table 1 presents the demographic characteristics and the pretest means and standard deviations for the treatment and control groups. The results of balance tests conducted on the demographic distributions and means on each of the preprogram spring literacy assessments (see Table 1, column 3) revealed that despite the point differences, the treatment and control group formed by random assignment did not statistically differ on any attribute or mean assessment score at the point at which treatment assignment occurred, $p > 0.05$.

[1] Three students (1 treatment, 2 control) included in the original random assignment pool ($N = 49$) transferred out of the district prior to the start of first grade and were excluded from all analyses.

Table 1
Student Demographics and Pretest Means and Standard Deviations by Assignment Status.

	Treatment		Control		
	M	*SD*	*M*	*SD*	M_D (SE_D)
Female	0.54	–	0.36	–	0.18(0.15)
Ethnic Minority	0.17	–	0.36	–	0.20(0.13)
English Learner	0.08	–	0.23	–	0.14(0.11)
Free Lunch Recipient	0.67	–	0.59	–	0.07(0.15)
Age in Months	66.42	3.45	65.23	3.24	1.19(0.99)
Spring Kindergarten LNF	34.25	7.57	35.77	8.17	1.52(2.32)
Spring Kindergarten NWF	31.96	9.38	34.95	7.88	2.99(2.57)
Spring Kindergarten PSF	49.33	12.68	56.50	13.18	7.17(3.82)
n	24		22		

Note. M_D = Mean Difference; SE_D = Standard Error of the Difference; LNF = Letter Naming Fluency, NWF = Nonsense Word Fluency, PSF = Phoneme Segmentation Fluency.

2.4. Outcome measures

At the time that study was conducted, the district administered a series of grade-specific assessments to monitor student learning progress each academic year. In kindergarten, the district utilized the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS is a standardized, individually administered measure of early literacy development that is appropriate for students for whom learning to read in English is a goal (Good & Kaminski, 2002). DIBELS contains multiple subtests, including Letter Naming Fluency (LNF), Phonemic Segmentation Fluency (PSF), and Nonsense Word Fluency (NWF) that are designed to track the pre-reading and early reading skill development of students (Good & Kaminski, 2002). Student performance is gauged by the number of correct responses given to an item battery (e.g., a list of letters or spoken phonemes) during one minute assessment frames. Students who perform with speed and precision earn higher subtest scores.

2.4.1. Letter naming fluency (LNF)

DIBELS' LNF subtest is a one minute assessment designed to measure students' ability to identify individual letters, a fundamental component of alphabetic knowledge (i.e., that printed letters symbolize and represent the sounds in spoken words). At each administration, students are presented with a page of 110 upper- and lower-case letters randomly arranged in 11 rows with 10 letters on each row. Students receive one point for each correct letter name provided during the assessment frame, with the total number of points serving as the LNF score. Letter naming has been established as a critical early predictor of future reading success (Ritchey & Speece, 2006; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004) and serves as an important risk factor in the development of reading disabilities (Catts, Fey, Zhang, & Tomblin, 2001; Puolakanaho et al., 2007). The test developer reports single (0.74) and multi-probe (0.90) reliability estimates (Good et al., 2004) and a one month alternative form reliability estimate of 0.93 (Good, Kaminski, Simmons, & Kame'enui, 2001). A two week test-retest reliability estimate of 0.90 has also been reported (Elliott, Lee, & Tollefson, 2001). Positive concurrent and predictive relationships between kindergarten LNF scores and a variety of curriculum-based measures (Burke, Hagan-Burke, Kwok, & Parker, 2009), and TerraNova vocabulary ($r = 0.63$), language ($r = 0.57$), and reading ($r = 0.48$) scores have also been reported (Rouse & Fantuzzo, 2006).

2.4.2. Nonsense word fluency (NWF)

DIBELS' NWF subtest is a one minute assessment designed to measure students' understanding of the alphabetic principle. Students are presented with randomly ordered nonsense words (e.g., vaj, ov) 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 0.83 in one student sample and a multi-probe reliability estimate of 0.98 in a second (Good et al., 2004). Concurrent and predictive relationships with a variety of early reading measures have also been reported (Burke et al., 2009; Dynamic Measurement Group, 2008). For example, in a sample of scores obtained from kindergarten children from a large urban school district, kindergarten NWF scores concurrently correlated 0.36–0.56 with scores on subtests of the Test of Early Reading Ability and later correlated 0.50, 0.57, and 0.55 with scores obtained at the end of first grade on the TerraNova reading, vocabulary, and language subtests, respectively (Rouse & Fantuzzo, 2006). In another study, moderately strong positive correlations were obtained between kindergarten NWF scores and student scores on several curriculum-based measures (Ritchey, 2008).

2.4.3. Phoneme segmentation fluency (PSF)

DIBELS' PSF subtest is designed to index a child's ability to fluently produce the individual sounds contained within a word. The ability to blend and segment phonemes is one of the strongest predictors of how well children learn to read over the early elementary school years (Ehri et al., 2001; National Reading Panel, 2000; Schatschneider et al., 2004). The PSF score is based on the number of phonemes correctly produced by students in response to a series of spoken three and four-phoneme words during a one minute assessment frame. The test developer reports single and multi-probe reliability estimates of 0.74 and 0.90, respectively (Good et al., 2004). Concurrent relationships with a variety of phonemic awareness indices and early reading measures have been identified (Burke et al., 2009; Dynamic Measurement Group, 2008; Hintze, Ryan, & Stoner, 2003; Rouse & Fantuzzo, 2006). The test developer also reports that student scores on the spring of kindergarten PSF assessment correlated 0.68 with their spring of first grade scores on the Woodcock-Johnson Psycho-Educational Battery Readiness Cluster (Good et al., 2004).

2.5. Administrative procedures and assignment criteria

DIBELS subtests were administered two to three times a year, each over a one week testing window. The first assessment occurred during the first week of classes in September, the second occurred directly after the winter break in January, and the third took place in mid-May, approximately four weeks prior to the end of the academic year. In kindergarten, the DIBELS' LNF subtest was administered at all three assessment occasions while the PSF and NWF subtests were administered in January and May. The LNF, PSF, and NWF subtests were also administered during the first week of classes in September of first grade. Summer changes in learning were measured by computing the difference in LNF, PSF, and NWF performance between the first assessment in first grade (mid September) and the final assessment in kindergarten (mid May).

Scores on the spring administration of DIBELS' NWF subtest were used to determine students' eligibility for a summer school placement. In addition to invitations distributed to those most at risk, summer school placements were also randomly assigned to a pool of kindergarten students ($N = 46$) who were projected to score above the cutpoint (NWF < 25) the district historically used to identify students in need of supplemental summer instruction.[2] The provision of summer school to students at moderately risk of future reading difficulty was made possible through a collaborative university-school district partnership that provided additional resources to expand the program to a wider range of students. Resources included additional funding for district and instructional staff, and graduate student support for observation and data collection.

[2] All students who scored below the cutpoint (NWF < 25) also received an invitation to attend summer school.

2.6. Analytic procedures

A multivariate analysis of variance (MANOVA) was performed with the summer change in DIBELS' LNF, PSF, and NWF scores as the dependent variables and treatment assignment status as the independent variable. Assignment status had two levels, treatment and control. The multivariate analysis was designed to align with the instructional targets (i.e., the development of multiple early literacy skills) of the summer literacy intervention and to contrast the performance of students exposed to multi-component literacy instruction with that of their peers not in receipt of targeted intervention. More specifically, the MANOVA analysis was designed to contrast the summer performance of those assigned to treatment (whether they participated or not) with the summer performance of the control group, where the outcome was a multivariate linear composite of key early literacy indicators. The intent-to-treat (ITT) analysis was thus used to provide a policy-relevant causal estimate of the treatment offer.

After conducting the MANOVA, supplemental analyses were performed to evaluate the impact of covariates on the ITT estimate, and to further isolate the effect of summer school for the subgroup of students who complied with their assignment (Angrist, Imbens, & Rubin, 1996; Imbens & Rubin, 1997; Little & Yau, 1998; Sagarin et al., 2014). In order to investigate the potential for covariate attenuation on the ITT estimate, discriminant function scores obtained from the MANOVA were regressed on the treatment assignment indicator, student demographic variables, and pretest scores. Weighting methods were also applied to the ITT estimate to take into account the impact of "no shows" (i.e., students who were assigned to treatment but refused to participate in summer school). By weighting the ITT estimate by the proportion of summer school invitees who actually attended ($M_{\text{assigned}} - M_{\text{control}}/\text{p}_\text{c}$, where p_c = the proportion of compliers), a "local" treatment effect estimate specific to the subgroup of summer school compliers was generated (Bloom, 1984; Gennetian, Morris, Bos, & Bloom, 2005). To compare the results from the manual weighting method and also allow for control of covariates, a two-stage least squares regression model was also estimated.

3. Results

Wilks' test of multivariate significance revealed that assignment status was statistically related to the weighted multivariate combination of early literacy measures, $\Lambda = 0.77$, F (3, 42) = 4.18, $p < 0.05$, $\eta^2 = 0.23$. The standardized discriminant function coefficients (SDFC) used to weight the multivariate composite revealed that PSF change (*SDFC* = 0.77), and NWF change (*SDFC* = 0.55) provided the largest independent contributions in the formation of the function that discriminated the groups. LNF change contributed less to the function (*SDFC* = 0.26). The structure coefficients demonstrated that the observed measures had moderate to strong correlations with the multivariate composite, PSF change ($r = 0.74$), NWF change ($r = 0.58$), LNF change ($r = 0.42$). Computation of the parallel discriminant ratio coefficients (DRC) revealed that PSF change (*DRC* = 0.77 * 0.74 = 0.57) was more important in distinguishing the groups than NWF change (*DRC* = 0.55 * 0.58 = 0.32) or LNF change (*DRC* = 0.26 * 0.42 = 0.11). On the basis of the statistically significant multivariate test, it can be concluded that the centroid associated with the treatment group was statistically larger ($M = 0.51$, $SD = 0.70$) than the centroid associated with the control group ($M = -0.56$, $SD = 1.25$), a 1.07 standardized unit difference.

The application of a correction factor for noncompliance with treatment assignment resulted in an upward adjustment to the ITT estimate. Weighting the ITT estimate by the proportion of treatment group members who accepted the treatment offer and actually received school-based summer instruction, the compliance-adjusted summer school estimate was approximately 2 linear composite units (i.e., 0.51 to −0.56 = 1.07, 1.07/.54 = 1.98). A two-stage least squares regression analysis using treatment assignment status as an instrumental variable also provided a direct estimate of the effect of summer school participation. Stage 1 estimated the relationship between treatment assignment and participation status. Stage 2 estimated the relationship between predicted participation status and summer literacy change. Results mirrored those obtained from applying Bloom's adjustment and indicated that the local effect of summer school was statistically significant at a conventional alpha level, $b = 1.98$, $SE = 0.59$, $p < 0.05$.

With regard to the individual measures that comprised the multivariate composite, the computation of univariate ANOVAs revealed a statistically significant mean difference between treatment groups on the summer change in PSF, $F(1, 44) = 7.18$, $MSE = 148.91$, $p < 0.05$. The treatment group experienced a mean gain ($M = 3.34$, $SD = 9.67$) while the control group demonstrated a mean performance loss ($M = -6.32$, $SD = 14.48$), Hedges' $g = 0.78$. The treatment group ($M = 4.17$, $SD = 16.12$) also outperformed the control group ($M = -4.68$, $SD = 12.01$) on the summer change in NWF, and the summer change in LNF (treatment group: $M = 4.04$, $SD = 9.08$); control group: ($M = -0.32$, $SD = 10.50$), but the differences were not statistically significant when an adjustment was applied (i.e., $\alpha = 0.05/3 = 0.017$) to maintain the probability of type I error at 0.05. When adjusted for non-compliance, the between group difference in summer PSF change (3.34 to −6.32/.54 = 17.88) was approximately 18 phonemes. Although the omnibus tests of the other two outcome variables were not statistically significant at the adjusted alpha level, the estimated between group difference for compliers was over 16 nonsense words (i.e., NWF change, 4.17 to −4.68/.54 = 16.39) while the summer change in LNF (4.04 to −0.32/.54 = 8.07) was 8 letter names.

To investigate the impact of covariates on the multivariate ITT estimate, standardized discriminant function scores obtained from the MANOVA analysis (Enders, 2003) were regressed on the treatment assignment indicator, pretest scores, and dichotomously coded demographic variables. The coefficients presented in Table 2 reveal that in the no covariate model, the size and statistical significance of the ITT treatment effect, as expected, was identical to the centroid difference obtained from the MANOVA analysis ($b = 1.07$). When demographic covariates were added to the model, the size of the treatment effect was similar to the no covariate model ($b = 1.16$, $SE = 0.32$), as none of the student demographic variables were related to scores on the discriminate function. In contrast, the addition of the individual pretests reduced the treatment effect estimate ($b = 0.85$, $SE = 0.29$), as performance on the NWF and PSF subtests were predictive of summer changes on the multivariate composite of early literacy indicators. Similarly, the compliance-adjusted estimate was also attenuated when covariates were added to the two-stage least squares regression model, $b = 1.58$, $SE = 0.54$, given the same pattern of association between covariates and the outcome observed in the original ITT analysis.

Table 2
Treatment Impact Model Results.

	Model 1		Model 2		Model 3	
	b	*SE*	*b*	*SE*	*b*	*SE*
Assigned to Tx	1.07*	0.30	1.16*	0.32	0.85*	0.29
Female			−0.43	0.32	−0.26	0.28
Ethnic Minority			−0.27	0.39	0.25	0.36
Free Lunch Recipient			−0.18	0.32	−0.28	0.28
English Learner			−0.05	0.49	−0.38	0.43
Age in Months			−0.05	0.05	−0.07	0.04
Spring Kindergarten LNF					0.01	0.02
Spring Kindergarten NWF					−0.05*	0.02
Spring Kindergarten PSF					−0.03*	0.01
R^2	0.23		0.31		0.53	

Note. Discriminant function scores obtained from the MANOVA analysis were regressed on the treatment assignment and control variables.
* $p < 0.05$.

4. Discussion

Summer holds the potential for students to explore new interests, to engage in activities that nurture creativity, and to spend additional time with family and friends. Yet for some students, the wealth of educational and instructional resources available during the academic year becomes more limited, restricting opportunities to maintain or improve on prior achievement gains (Pallas, 2016; Slates et al., 2012). For students who stagnate or lose ground over the summer, increasingly difficult challenges often await when they reengage with the academic environment in the following school year. With respect to the development of key early literacy components, students who experience a summer fallback in learning may become frustrated with more demanding school-based reading activities, avoid opportunities to read, and therefore fail to engage in practices that develop reading skill. As they mature, initial reading deficiencies may increase and begin to impact performance in comprehension-dependent subject areas, further depressing academic success (Stanovich, 1986, 1992, 2000). The negative feedback loop, if left unchecked, eventually can translate into more far reaching negative academic outcomes, including lower grades, higher absenteeism, and greater risk of dropping out before degree completion (Cunningham & Stanovich, 1997; Reschly, 2010).

To stem the downward spiral, early and active intervention efforts in beginning reading are now viewed as the primary way to prevent initial reading difficulties from progressing into long-term reading failure (National Early Literacy Panel, 2008; National Reading Panel, 2000). In particular, the use of proactive screening and daily, explicit training in core literacy skills with a focus on teacher modeling of skills, frequent opportunity for guided and independent student practice coupled with individualized feedback has been shown to facilitate the development of early reading skills (e.g., Hurry & Sylva, 2007; Partanen & Siegel, 2014; Vaughn et al., 2009; Vellutino, Scanlon, Small, & Fanuele, 2006; Wanzek et al., 2016). The provision of a robust summer school program using best practice reading instruction is thus recommended for keeping academic resources flowing toward students most in need of learning support over the summer months (Kim & Quinn, 2013). However, due to the nonrandom nature in which students are often recruited for a summer school placement, it is often challenging to obtain unbiased program effect estimates (Cooper et al., 2000; Stein & Fonseca, 2016). The current study was thus designed to take advantage of a unique opportunity for evaluating the efficacy of a school-based summer literacy intervention. Through a collaborative partnership formed between university researchers and school district personnel, the effect of assignment to, and participation in, summer school could be evaluated.

Results of the multivariate ITT analysis indicated that students assigned to summer school outperformed their peers by approximately 1 standard deviation on a linear composite of literacy indicators during the summer between kindergarten and grade 1. Of the literacy outcomes that formed the function that discriminated the groups, the change in PSF was most consequential ($DRC = 0.57$). However, it should be reiterated that not all students who were invited to summer school accepted the offer. Application of an adjustment to evaluate the effect of compliance for the subgroup of students who were assigned and participated in summer school yielded a larger treatment impact estimate. Weighting the ITT estimate by the proportion of compliers, the "local" effect of summer school was approximately twice as large as that observed in the original analysis.

On the individual measures that formed the literacy composite, the unadjusted group difference in summer change outcomes ranged from 4 letter names to 9 phonemes and 9.5 nonsense words. Notably, students not assigned to summer school began the next academic year at a lower performance level than where they ended the previous year on each measure. In contrast, students assigned to summer school demonstrated uniformly higher performance. From a practical standpoint, the summer gains of students in the treatment group were large enough on average to move these students to the low risk (LNF > 37, NWF > 24) or established (PSF > 35) proficiency categories at the beginning of first grade. Overall, the summer literacy gain experienced by summer school participants and the literacy loss experienced by their non-treated peers was consistent with meta-analytic findings documenting summer school effects and the summer learning outcomes of children who do not attend a school-based summer program (Cooper et al., 2000; Cooper et al., 1996; Kim and Quinn, 2013).

The pattern of results was also largely consistent with findings associated with other randomized field trials (Borman, Goetz, & Dowling, 2009; Schacter & Jo, 2005) and strong quasi-experimental investigations that have utilized regression discontinuity designs (Jacob & Lefgren, 2004; Matsudaira, 2008) to examine the effects of supplemental summer instruction in elementary school student samples. Similar to current findings, summer school students in these studies gained literacy skills over the summer and/or demonstrated higher proficiency on a key literacy measure (relative to their control-group peers) at the beginning of the following academic year. Further, as with the results associated with another randomized field experiment that adjusted for non-compliance among low income students assigned to summer school (Borman & Dowling, 2006), a larger treatment effect was obtained after using statistical methods to adjust for non-compliance with the treatment offer. However, it is also important to note that results from a recent large scale multisite randomized trial contrast with the findings reported here and elsewhere. Across five urban districts, while the short term effect of providing 5 weeks of summer school to at risk students was statistically significant in mathematics, results in reading, although positive, were not large enough to be statistically or practically significant (McCombs et al., 2014). Moreover, when a longer term follow-up was conducted on a variety of achievement indicators over the next two academic years, no statistical differences between treatment and control students were observed, perhaps owing to high non-compliance rates during the summer intervention period (Augustine et al., 2016).

In addition to the contrast in results with the urban multisite study, a difference between the findings of this study and others where positive summer school impacts have been observed centers on the size of the treatment effect. With the unadjusted group contrast exceeding 1 standard deviation, we necessarily need to consider the design and sample characteristics that distinguished our investigation. First, it should be noted that within an intervention context that facilitated a high fidelity and low variance delivery of scripted, direct literacy instruction by experienced teachers (Zvoch, 2012), students received a strong dose of the intended treatment. The analytic sample also consisted of students who were moderately at-risk of developing future reading difficulties and thus may have been more able to respond to and benefit from the supplemental instruction (see Cooper et al., 2000). When combined with the use of assessments sensitive to the literacy skill development that was the primary focus of the instructional model, the reported effects were larger than those observed in studies that have investigated more diffuse, multi-component summer programs with more distal outcome measures (e.g., Borman & Dowling, 2006; Kim & Quinn, 2013; McCombs et al., 2014). As a result, caution is urged in generalizing current findings to dissimilar samples and intervention contexts. Additional study with larger, more representative samples and inclusion of more comprehensive measures is recommended to more clearly identify the conditions and the extent to which summer instruction can serve to promote summer learning gains.

As the provision of targeted summer instruction has been shown to offset the summer slide and promote achievement gains for students at risk of negative summer learning outcomes (Cooper et al., 2000; Kim & Quinn, 2013; McCombs et al., 2011; Zvoch & Stevens, 2015), school administrators may view summer school as a cost-effective means for delivering supplemental resources to those most in need (McCombs et al., 2014).. However, as summer programs vary by provider type (i.e., whether the program is offered by the district or an external entity), dosage and content (i.e., length, strength, and subject matter), setting (i.e., whether the program is site-based or home-based), and purpose (i.e., remediation, acceleration, recreation), intervention outcomes have to be considered with respect to context in which they were achieved (McCombs et al., 2011). In the current study, an intense dose of summer instruction delivered by highly trained and experienced teachers to students at moderate risk generated strong positive results on a curriculum-aligned assessment. In other contexts where the content, instructional method, outcomes, and age and risk profile of students differs, the effect may vary by student group or intervention site or may not be as pronounced in particular academic or behavioral domains (Augustine et al., 2016; McCombs et al., 2014).

The investigation of the size and direction of summer school effects also requires attention to the manner in which students are assigned a summer school placement. For defensible inferences to be drawn, a strong research design involving random assignment should be used when possible (Borman & Dowling, 2006; Borman et al., 2005). When random assignment is not ethical or feasible, less disruptive methods such as the regression discontinuity and interrupted time series design may offer a practical alternative for separating instructional effects from the background characteristics of students (Murnane & Willett, 2011; Shadish et al., 2002). Within the context of a summer school intervention, it is also important to monitor participation in alternative summer enrichment activities and the frequency and amount of independent summer reading completed by treatment and control students. The collection of data on the summer learning activities of all study participants permits researchers to document the “business-as-usual” counterfactual condition and more thoroughly contextualize the nature of the treatment contrast (Weiss, Bloom, & Brock, 2014).

As summer school attendance is often voluntary, it is also recommended that researchers carefully consider how to approach breakdowns in treatment compliance as the nature and generalizability of the inference regarding program performance can vary in relation to the intended and achieved design and the analytic model applied to data. In this regard, the collection and analysis of data on the summer educational opportunities and experiences of students may hold value in identifying the reasons students (and parents) accept or decline an invitation to summer school (Borman et al., 2005). To the extent that researchers can elucidate the conditions that promote summer learning engagement and obtain an inferentially strong and contextually grounded treatment impact estimate, school leaders will be better equipped to make informed decisions regarding the best approaches for delivering effective, targeted summer instruction to those with the greatest need.

Acknowledgement

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

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.

Atteberry, A., & McEachin, A. (2016). School's out: Summer learning loss across grade

levels and school contexts in the United States today. In K. A. Alexander, S. Pitcock, & M. Boulay (Eds.). *The summer slide: What we know and can do about summer learning loss* (pp. 35–54). New York, NY: Teachers College Press.
Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J., McEachin, A., et al. (2016). *Effects of voluntary summer learning programs on low-income urban youth.* Santa Monica, CA: RAND Corporation.
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., & 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., Benson, J., & Overman, L. T. (2005). Families, schools, and summer learning. *Elementary School Journal, 106*(2), 131–150.
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.
Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy indicators from the middle of kindergarten to second grade. *The Journal of Special Education, 42*(4), 209–226.
Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*, 38–50.
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: Lawrence Erlbaum Associates.
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., 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., 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.
Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945.
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.
Dynamic Measurement Group (2008). *DIBELS 6th edition technical adequacy information (Technical report No. 6)*Eugene, OR: Author. Available: http://dibels.org/pubs.html.
Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250–287.
Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the dynamic indicators for basic early literacy skills–Modified. *School Psychology Review, 30*(1), 33–49.
Enders, C. K. (2003). Performing multivariate group comparisons following a statistically significant MANOVA. *Measurement and Evaluation in Counseling and Development, 36*, 40–56.
Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality.* Boulder, CO: Westview Press.
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.
Gershenson, S. (2013). Do summer time-use gaps vary by socioeconomic status? *American Educational Research Journal, 50*(6), 1219–1248.
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. (NCEE 2009-4045).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides.
Good, R. H., & 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., Simmons, D., & Kame'enui, E. J. (2001). *Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model: Steps to reading outcomes.* Eugene, OR: University of Oregon.
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.
Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test of phonological processing. *School Psychology Review, 32*(4), 541–556.
Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading, 30*(3), 227–248.
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.
Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics, 86*(1), 226–244.
Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research, 83*(3), 386–431.
Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading, 12*(1), 1–23.
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.
Matsudaira, J. D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics, 142*, 829–850.
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.
McCombs, J. S., Pane, J. F., Augustine, C. H., Schwartz, H. L., Martorell, P., & Zakaras, L. (2014). *Ready for fall? Near-term effects of voluntary summer learning programs on low-income students' learning opportunities and outcomes.* Santa Monica, CA: RAND Corporation.
Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research.* New York, NY: Oxford University Press.
National Association for Year-Round Education (2007). *Statistical summaries of year-round education programs, 2006–2007.* San Diego, CA: Author.
National Early Literacy Panel (2008). *Developing early literacy.* Washington, DC: National Institute for Literacy.
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).
Pallas, A. M. (2016). The sunny home advantage: Class differences in resources to support summer learning. In K. A. Alexander, S. Pitcock, & M. Boulay (Eds.). *The summer slide: What we know and can do about summer learning loss* (pp. 111–130). New York, NY: Teachers College Press.
Partanen, M., & Siegel, L. S. (2014). Long-term outcomes of the early identification and intervention of reading disabilities. *Reading & Writing, 27*(4), 665–684.
Patall, E. A., Cooper, H., & Batts Allen, A. (2010). Extending the school day or school year: A systematic review of research (1985-2009). *Review of Educational Research, 80*(3), 401–436.
Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppnen, P. H. T., Poikkeus, A. M., et al. (2007). Very early phonological and language skills: Estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry, 48*(9), 923–931.
Rambo-Hernandez, K. E., & McCoach, D. B. (2015). High-achieving and average students' reading growth: Contrasting school and summer trajectories. *Journal of Educational Research, 108*(2), 112–129.
Reschly, A. L. (2010). Reading and school completion: Connections and Matthew effects. *Reading & Writing Quarterly, 26*(1), 67–90.
Ritchey, K. D., & Speece, D. L. (2006). From letter names to word reading: The nascent role of sublexical fluency. *Contemporary Educational Psychology, 31*, 301–327.
Ritchey, K. D. (2008). Assessing letter sound knowledge: A comparison of letter sound fluency and nonsense word fluency. *Exceptional Children, 74*(4), 487–506.
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.
Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological Methods, 19*(3), 317–333.
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., Fletcher, J. M., Francis, D. J., Carlson, C. D., & 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.
Skinner, R. R. (2014). *Year-round schools: In brief.* Washington, DC: Congressional Research Service.
Slates, S. L., Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2012). Counteracting summer slide: Social capital resources within socioeconomically disadvantaged families. *Journal of Education for Students Placed At Risk, 17*(3), 165–185.
Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–406.
Stanovich, K. E. (1992). *Speculations on the causes and consequences of individual differences in early reading acquisition.* Hillsdale, NJ, England: Lawrence Erlbaum Associates.
Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York: Guilford Publications.
Stein, M. L., & Fonseca, E. (2016). A review and analysis of evaluations of summer programs. In K. A. Alexander, S. Pitcock, & M. Boulay (Eds.). *The summer slide: What we*

know and can do about summer learning loss (pp. 271–283). New York, NY: Teachers College Press.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*, 7–26.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention examining higher and lower responders. *Exceptional Children, 75*(2), 165–183.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39*(2), 157–169.

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*.

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.

von Hippel, P. T. (2016). Year-round school calendars: Effects on summer learning, achievement, parents, teachers, and property values. In K. A. Alexander, S. Pitcock, & M. Boulay (Eds.). *The summer slide: What we know and can do about summer learning loss* (pp. 208–230). New York, NY: Teachers College Press.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review, 28*(3), 551–576.

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management, 33*(3), 778–808.

Zvoch, K. (2012). 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, 33*(4), 541–559.