

# Can a Mixed-Method Literacy Intervention Improve the Reading Achievement of Low-Performing Elementary School Students in an After-School Program? Results From a Randomized Controlled Trial of READ 180 Enterprise

**James S. Kim**

**Lauren Capotosto**

*Harvard University*

**Ardice Hartry**

*University of California, Berkeley*

**Robert Fitzgerald**

*MPR Associates, Inc.*

*The authors describe an independent evaluation of the READ 180 Enterprise intervention designed by Scholastic, Inc. Despite widespread use of the program with upper elementary through high school students, there is limited empirical evidence to support its effectiveness. In this randomized controlled trial involving 312 students enrolled in an after-school program, the authors generated intention-to-treat and treatment-on-the-treated estimates of the program's impact on several literacy outcomes of fourth, fifth, and sixth graders reading below proficiency on a state assessment at baseline. READ 180 Enterprise students outperformed control group students on vocabulary ( $d = .23$ ) and reading comprehension ( $d = .32$ ) but not on spelling and oral reading fluency. The authors interpret the findings in light of the theory of instruction underpinning the READ 180 Enterprise intervention.*

*Keywords: reading intervention, adolescent literacy, after-school programs, reading difficulties, randomized experiments*

THE reading ability of U.S. adolescents has been an enduring concern among researchers and policy makers. Although a higher percentage of fourth graders read at the proficient level on the National Assessment of Educational Progress in 2007 than in previous years, there has been no change in the percentage of eighth graders reading at or above this level since 1992. Approximately 70%

of eighth graders have consistently scored below proficiency (Lee, Grigg, & Donahue, 2007). In his analysis of National Assessment of Educational Progress data, Loveless (2007) noted that reading achievement from fourth grade to eighth grade “is languishing if not deteriorating” (p. 8).

To address this “adolescent literacy crisis” (Carnegie Council on Advancing Adolescent

Literacy, 2010; Deschler, Palinscar, Biancarosa, & Nair, 2007), many public school districts have turned to commercially developed literacy interventions such as READ 180, a Scholastic, Inc., product. READ 180 uses a mixed-method approach to literacy instruction among struggling readers in the upper elementary and middle grades (Slavin, Cheung, Groff, & Lake, 2008). Through a combination of teacher-directed instruction, computer-based reading lessons, and independent reading, the program targets upper elementary through high school students who score below proficiency on state performance assessments (Hasselbring & Goin, 2004). With Chicago Public Schools expanding its READ 180 program to serve 8,600 students in 80 schools in 2009–2010, there is evidence of the intervention's growing demand in some districts (Chicago Public Schools, 2009).

Scholastic makes substantial claims about READ 180's effectiveness. It asserts to be appropriate for the most struggling readers and recommends that schools target students in the 25th percentile, or with stanines of 1 to 3 on standardized reading assessments (Scholastic, Inc., n.d.). READ 180 is marketed as a research-based program that can improve reading levels of participating students by 2 to 5 years by addressing the many different aspects of literacy: phonemic and phonological awareness, fluency, vocabulary, reading comprehension, spelling, and writing. Despite claims regarding the efficacy of READ 180 and its widespread use in public school classrooms, there is surprisingly little experimental evidence backing the claims of the developers (Slavin et al., 2008; What Works Clearinghouse, 2009).

### *Overview of Study*

The purpose of this randomized controlled trial is to evaluate the efficacy of READ 180 Enterprise on measures of vocabulary, reading comprehension, spelling, and oral reading fluency among students in Grades 4 to 6, 95% of whom scored below proficiency on the Massachusetts Comprehensive Assessment System (MCAS) in English language arts. We also test the efficacy of READ 180 Enterprise in an after-school setting to examine whether a more structured literacy program can generate

larger gains than a less structured enrichment program with little formal literacy instruction. Students in both groups attended their after-school program 4 days per week (2 hours per day) over 23 weeks from October 2006 to April 2007. By controlling the amount of time children spent in the program, we isolated the effects of two different instructional interventions. The remainder of the introduction describes some reasons why adolescents struggle to read with comprehension and the READ 180 program theory and evidence base. Next, we turn to the study methods and results and conclude with a discussion of the findings.

### *Why Do Adolescents Struggle to Read for Understanding?*

Multiple factors can account for adolescent reading difficulties. The Simple View of Reading (SVR; Gough & Tunmer, 1986; Hoover & Gough, 1990) suggests that reading comprehension is the product of word recognition and language or listening comprehension. Given the multiplicative nature of the SVR, breakdowns in either component can affect reading comprehension. By the upper elementary and middle grades, there is also significant heterogeneity in the reading profiles of struggling readers. For example, Buly and Valencia (2002) identified 10 subgroups of fourth grade readers who were below proficient on a Washington State reading assessment. Although some students in their sample struggled with decoding and word recognition, the majority had a range of difficulties, including limited vocabularies.

Without intensive intervention, children with reading difficulties are at risk for falling increasingly behind their more reading proficient peers in the upper elementary and middle grades (Chall, Jacobs, & Baldwin, 1990; Juel, 1988). This gap-widening phenomenon, referred to as the "Matthew effect" (Stanovich, 1986), has been attributed in part to differences in reading volume (Cunningham & Stanovich, 1998). Students with reading difficulties are often less motivated to engage in self-initiated reading activities than their peers who are proficient word readers (Snow, Burns, & Griffin, 1998). With less exposure to print, poor readers have fewer opportunities than their peers to practice reading fluency,

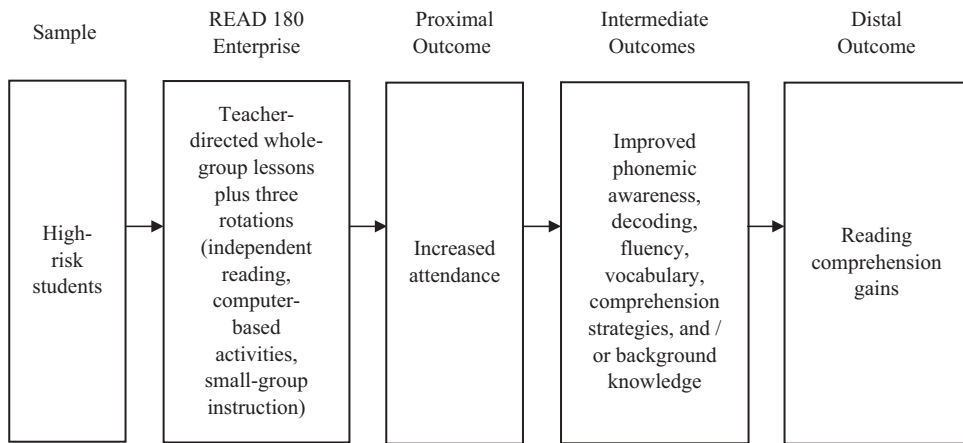


FIGURE 1. *READ 180 Enterprise logic model.*

encounter academic vocabulary, develop content-area knowledge, and interact with abstract ideas and complex writing structures (Chall et al., 1990; Cunningham & Stanovich, 1998; Kieffer & Lesaux, 2007; Stanovich, 1986). Remedial reading interventions in the upper elementary grades must address the large gaps in reading practice and print exposure that underlie the gap in reading skill between good and poor readers (Torgesen, 2005).

Because poor readers often require intensive interventions to compensate for years of underachievement, programs that supplement the regular English language arts curriculum may provide the additional time poor readers need to improve their achievement (U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, 2007). These programs may have the greatest impact when the curriculum is structured and focused on addressing the literacy skills of struggling readers. However, results from several recent evaluations and research syntheses indicate that after-school programs typically serve multiple goals and do not consistently improve reading achievement (Granger, 2008; Kane, 2004; Lauer et al., 2006; Zief, Lauver, & Maynard, 2006). For example, the national evaluation of the 21st Century Community Learning Centers Program, which provides students with a mix of academic, social, and recreational activities, showed no main effect on reading achievement among a sample of elementary

(Grades 1 to 6) students who scored below the national norm on the Stanford Achievement Test (SAT) 9 reading test (James-Burdumy, Dynarski, & Deke, 2007). Further study is needed to test the hypothesis that a highly structured and engaging literacy curriculum in an after-school setting would improve the reading gains of less proficient readers in the upper elementary grades.

#### *READ 180: Program Theory and the Evidence Base*

READ 180 aims to address the diversity in student reading profiles by providing differentiated instruction in each of the components of reading: phonemic and phonological awareness, fluency, vocabulary, and reading comprehension. According to the READ 180 logic model, outlined in Figure 1, the efficacy of the program depends on targeting children scoring below the 25th percentile (i.e., high-risk readers) through whole-group teacher-directed lessons and the three rotations: individualized computer-assisted reading instruction, independent and modeled reading practice with leveled text, and teacher-directed reading lessons tailored to the reading level of small groups. The logic model provides a heuristic for understanding the program theory and the conditions under which the READ 180 intervention is likely to improve reading comprehension (Rossi, Lipsey, & Freeman, 2004).

The instructional activities of the program have support in the reading literature. In its full 90-minute version, READ 180 offers whole-group teacher-directed instruction (20-minute introduction and 10-minute wrap-up), during which teachers build background and activate students' prior knowledge (Butcher & Kintsch, 2003) and model fluency and comprehension strategies (Palinscar & Brown, 1984). Individualized computer-assisted reading instruction (20 minutes) offers students opportunities to develop the critical reading components of decoding, fluency, vocabulary, and comprehension skills (National Institute of Child Health and Human Development, 2000). Independent and modeled reading practice with leveled text (20 minutes) aims to increase reading motivation and subsequent time spent reading through self-selected high-interest books and audio books (Guthrie, 2003). Teacher-directed reading lessons tailored to the reading level of small groups (20 minutes) allows teachers to differentiate instruction (Tomlinson, 2001), a critical feature given the heterogeneity of student reading profiles (Buly & Valencia, 2002; Morris et al., 1998).

Although Scholastic claims that 37 studies have proven the effectiveness of READ 180 on student achievement (Scholastic Research, 2008), this research base is limited by three factors. First, over a quarter of these studies were designed and published by Scholastic. Such developer-designed studies tend to yield larger effects than independent evaluations (Borman, Hewes, Overman, & Brown, 2003). Second, the vast majority of READ 180 studies have used quasi-experimental methods (Slavin et al., 2008; What Works Clearinghouse, 2009), which tend to yield larger effect sizes than more carefully controlled studies (Borenstein, Hedges, Higgins, & Rothstein, 2009). Finally, although READ 180 provides instruction in several components of reading, the majority of studies have measured only reading comprehension and general literacy achievement outcomes (What Works Clearinghouse, 2009). None of the studies that met What Works Clearinghouse standards examined the impact of the program in the word reading or fluency domains.

Given these limitations, researchers who have synthesized the READ 180 literature offer

tentative conclusions. Slavin et al. (2008) concluded that there was "moderate" (as opposed to "strong") evidence of a positive impact on reading comprehension ( $d = .24$ ), while a more recent What Works Clearinghouse (2009) assessment found a medium effect for comprehension (+4 percentile points) and large effect for general literacy achievement (+12 percentile points). The larger effects of this more recent review are driven in part by results from a developer-designed quasi-experimental study ( $d = .45$ ; Scholastic Research, 2008) that was not included in Slavin et al.'s best-evidence synthesis.

One recent randomized experiment (Lang et al., 2008) found differential effects of READ 180 on the basis of prior student reading ability. In their study of 1,265 struggling ninth grade readers, Lang et al. (2008, 2009) found a negative, nonsignificant effect ( $d = -.27$ ) of READ 180 on high-risk students reading below the fourth grade level on the Florida Comprehensive Assessment Test (FCAT) and a significant positive effect ( $d = .30$ ) of the program on moderate-risk students reading between the fourth and sixth grade levels. The corresponding average score on the SAT 9 was the 44th percentile for moderate-risk students (i.e., Level 2 of the FCAT), whereas the average for high-risk students was at the 25th percentile (i.e., Level 1 of the FCAT). This finding was particularly noteworthy given Scholastic's recommendation to focus specifically on this high-risk group.

Investigating the effects of the intervention as an after-school program on multiple reading outcomes, we (Kim, Samson, Fitzgerald, & Hartry, 2010) found differential effects by grade level. Nearly 300 fourth to sixth grade students who scored below proficiency on a state English assessment were randomly assigned to either READ 180 Version 1.6 or the district's regular after-school program. The READ 180 treatment was modified from its recommended 90-minute block to fit within a 60-minute period, and it included only the three rotations: individualized computer-assisted reading instruction, independent and modeled reading practice with leveled text, and teacher-directed reading lessons tailored to the reading level of small groups. No whole-group teacher-directed lessons were implemented. Although the findings

indicated a positive main effect of READ 180 on oral reading fluency ( $d = .12$ ) and attendance rates ( $d = .31$ ), these results were driven by a positive effect for fourth grade children only. There was no main effect, however, on measures of word reading efficiency, reading comprehension, and vocabulary. Because the students in this study scored at the 25th percentile on a standardized test of reading comprehension at the beginning of the study, the nonsignificant effects on the reading comprehension test from our earlier study are similar to the findings reported by Lang et al. (2008, 2009) for high-risk students who scored at the 25th percentile. Collectively, the most recent experimental studies suggest that READ 180 has differential effects on students of varying grades and reading achievement.

With three noteworthy differences, the current study builds on our previous work (Kim et al., 2010). First, students in the current sample were higher performing readers than students in our previous study. Second, whereas READ 180 students participated in only the program's three small-group rotations, READ 180 students in the current study also received whole-group teacher directed lessons. This additional focus reflects READ 180's upgrade from Version 1.6 to the Enterprise Edition. Third, whereas teachers in our previous study developed their own lesson plans for small-group instruction, teachers in the current study drew from the rBook Teacher's Edition, which provides detailed lesson plans to build vocabulary, comprehension, and fluency.

### *Study Goals*

Three study goals motivated our current study. First, we measure several reading outcomes that READ 180 Enterprise is designed to improve. Using an experimental design, we generate impact estimates on posttest measures of vocabulary, reading comprehension, spelling, and fluency. This approach aims to improve on most randomized controlled trials of READ 180, which measure impact estimates on a single measure of reading comprehension and do not permit a direct test of the theory of instruction underlying the intervention (Lang et al., 2009; Slavin et al., 2008). READ 180 Enterprise

targets multiple components of reading because the theory of instruction is that reading comprehension is the product of word reading and language comprehension (Gough & Tunmer, 1986) and that struggling readers have varying reading profiles (Buly & Valencia, 2002; Morris et al., 1998).

Second, we examine whether the impact of offering READ 180 Enterprise to upper elementary school children differs by grade level. Differential effects by grade may be observed because of differences in attendance. Research on after-school programs suggests that attendance rates are lower, on average, for middle school than elementary school-aged children (Kane, 2004; U.S. Department of Education, Office of the Undersecretary, 2003). Insofar as younger children attend a voluntary after-school program, they may reap larger benefits than older children who have lower participation rates.

Third, we estimate the impact of the treatment-on-the-treated (TOT) by using initial random assignment as an instrument for the number of days that a student attended the READ 180 Enterprise program. Because participation in after-school programs is voluntary, policy makers cannot force children to comply with attendance requirements. Nonetheless, identifying the effects of attending READ 180 Enterprise on student outcomes would provide important evidence on the impacts policy makers could anticipate if children actively attend and participate in the READ 180 Enterprise after-school program. To date, however, no studies have used instrumental variables to estimate the impact of attending READ 180 on reading outcomes (i.e., the TOT estimate; Angrist, Imbens, & Rubin, 1996).

Our study goals motivated the following research questions: (a) What is the intention-to-treat (ITT) estimate of READ 180 Enterprise in a voluntary after-school program designed to improve the reading achievement of low-performing children in the upper elementary grades? Does the impact of offering the READ 180 Enterprise after-school program differ by grade level? and (b) What is the TOT estimate of READ 180 Enterprise in a voluntary after-school program designed to improve the reading achievement of low-performing children in the upper elementary grades?

TABLE 1  
*Descriptive Characteristics of the Sample at the Beginning of the Study (n = 312)*

Variable	%
Grade	
4	36
5	44
6	20
Gender	
Female	54
Male	46
Eligible for free or reduced-price lunch	
No	31
Yes	69
Ethnicity	
White	28
African American	54
Latino/a	12
Other	6
MCAS 2006 (English language arts)	
Proficient/advanced	5
Needs improvement	80
Failing	15

*Note.* Massachusetts Comprehensive Assessment System (MCAS) 2006 performance levels were available for 295 students.

## Method

### *Sample and Measures*

Part of a larger implementation study (Hartry, Fitzgerald, & Porter, 2008), this study was undertaken in a mid-sized urban district in southeastern Massachusetts. Participating children in Grades 4 to 6 were recruited from four K–6 elementary schools where a majority of children were from low-income families and minority backgrounds. As shown in Table 1, 95% of the sample included children who scored below proficiency on the MCAS, a standards-based assessment of the state English language arts curriculum. Overall, the sample included three grade-specific cohorts of students who were failing to meet grade level expectations on the MCAS.

*SAT 10 reading comprehension, vocabulary, and spelling.* The SAT 10 is a standardized multiple-choice assessment, normed in 2002. The current study used the abbreviated battery, which contains a subset of items from the full battery. Across grades, Kuder-Richardson Formula 20

reliability coefficients for the abbreviated battery ranged from .76 to .82 in vocabulary and .84 to .86 in comprehension.

*Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency (DORF).* The DORF is a standardized, individually administered assessment of first through sixth grade students' reading accuracy and reading rate with connected text (Good & Kaminski, 2002). The number of words correctly read per minute is the oral reading fluency rate, and test–retest reliabilities range from .92 to .97.

*Attendance.* After-school teachers and supervisors kept student attendance records from the beginning of the study in October 2006 to the end of the study in April 2007.

*Fidelity of READ 180 implementation measures.* As part of a larger implementation study (for details, see Hartry et al., 2008), trained observers conducted three observations (November, February, and April) to assess the fidelity of READ 180 Enterprise implementation along seven intervention components. Each intervention component was rated using a 3-point scale (1 = low fidelity, 3 = high fidelity). There was moderate to high fidelity in classroom setup of the independent reading area ( $M = 2.64$ ,  $SD = 0.44$ ), computers ( $M = 2.65$ ,  $SD = 0.64$ ), and headsets ( $M = 2.71$ ,  $SD = 0.51$ ). Raters observed moderate to high fidelity of the small-group rotation ( $M = 1.94$ ,  $SD = 0.47$ ), computer rotation ( $M = 2.16$ ,  $SD = 0.50$ ), and independent reading rotation ( $M = 2.29$ ,  $SD = 0.61$ ). The lowest mean fidelity score was for whole-group lessons ( $M = 1.11$ ,  $SD = 0.68$ ) because the number of classrooms implementing teacher-directed whole-group lessons declined from 12 in November to 7 in April. In particular, the lower scores for fidelity of rotations resulted from teachers' decisions to either eliminate components of READ 180 Enterprise, such as whole-group wrap-up, or to minimize the number of minutes children spent in each rotation. Using scores on each of the seven intervention components, we created an overall mean fidelity score ( $M = 2.21$ ,  $SD = 0.30$ ) and coded whether each classroom's mean fidelity score was above or below the median of 2.235.

TABLE 2

*Comparison of Baseline Characteristics for Children in READ 180 Enterprise and Control Group*

Variable	Control			READ 180			<i>t</i>	<i>p</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Full sample								
White students	157	0.31	0.46	155	0.25	0.44	1.064	.288
Eligible for free/reduced-price lunch	157	0.71	0.46	155	0.67	0.47	0.686	.493
Female	157	0.56	0.50	155	0.52	0.50	0.671	.503
Age (months)	156	126.63	11.06	155	126.71	11.17	-0.064	.949
MCAS ELA proficient	150	0.05	0.23	145	0.06	0.23	-0.070	.945
DORF	157	88.03	26.98	155	88.72	27.33	-0.225	.822
Grade 4								
White students	57	0.33	0.48	55	0.29	0.46	0.480	.632
Eligible for free/reduced-price lunch	57	0.68	0.47	55	0.69	0.47	-0.076	.940
Female	57	0.61	0.49	55	0.45	0.50	1.699	.092
Age (months)	57	115.58	5.55	55	116.09	5.76	-0.482	.631
MCAS ELA proficient	56	0.05	0.23	53	0.06	0.23	-0.069	.945
DORF	57	77.16	25.33	55	75.51	22.48	0.364	.717
Grade 5								
White students	68	0.31	0.47	70	0.21	0.41	1.263	.209
Eligible for free/reduced-price lunch	68	0.71	0.46	70	0.70	0.46	0.075	.940
Female	68	0.41	0.50	70	0.53	0.50	-1.374	.172
Age (months)	67	129.40	5.95	70	128.52	6.47	0.823	.412
MCAS ELA proficient	63	0.06	0.25	67	0.04	0.21	0.469	.640
DORF	68	90.34	23.50	70	91.51	25.84	-0.279	.780
Grade 6								
White students	32	0.25	0.44	30	0.27	0.45	-0.148	.883
Eligible for free/reduced-price lunch	32	0.75	0.44	30	0.57	0.50	1.528	.132
Female	32	0.78	0.42	30	0.63	0.49	1.279	.206
Age (months)	32	140.52	6.35	30	141.94	6.30	-0.888	.378
MCAS ELA proficient	31	0.03	0.18	25	0.08	0.28	-0.779	.439
DORF	32	102.50	29.43	30	106.43	27.60	-0.542	.590

Note. DORF = Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency; ELA = English language arts;

### *Design for Random Assignment*

*Power analysis.* We conducted a power analysis to identify the number of children needed to detect an effect size of .20 on a standardized test of reading using a two-tailed test with  $\alpha$  set at .05. Our power analysis indicated that approximately 300 students were required to have an 80% chance of detecting effect sizes of .20 using a covariate that was correlated .70 to .80 with the posttest.

*Student recruitment plan.* In September and October 2006, students from four elementary schools were recruited to participate in this study. Given our target sample size, we initially recruited Grade 4 to 6 students who scored below proficient on the

Massachusetts Comprehensive Assessment System (MCAS) English language arts assessment in spring 2006. A small number of students scoring above proficient (5% of the sample) were recruited to reach our target sample. Students who returned active consent forms were administered baseline assessments and included in the random assignment protocol, which took place 2 weeks before the beginning of the after-school program.

To improve the precision of the treatment effects, we stratified children by school and grade, which we refer to as randomization blocks, and then randomly assigned children to either READ 180 Enterprise or the district after-school program. As shown in Table 2, the comparison of READ 180 Enterprise and control children at baseline revealed no statistically significant

difference on baseline characteristics for the full sample and by grade level. Differential attrition also did not threaten the internal validity of the findings. In particular, there was no statistically significant relationship between condition and the percentage of children who took the SAT 10 reading vocabulary and comprehension test,  $\chi^2(1, 312) = 2.291, p = .13$ , or the SAT 10 spelling test,  $\chi^2(1, 312) = 1.488, p = .22$ . Because posttests were administered to all students who remained in the school district and during the school day, most children in the study completed the reading posttests. For example, reading vocabulary and comprehension scores were available for 97% of the READ 180 Enterprise students ( $n = 150$ ) and 93% of the control group ( $n = 146$ ). Students who were not tested were lost to attrition outside of the district.

There was also no evidence that differential attrition of particularly high- or low-performing students affected the composition of the final analytic sample relative to the baseline sample. Among READ 180 Enterprise students, the pretest oral reading fluency score for the children who remained in the study ( $M = 88.49, SD = 26.45$ ) at posttest and those who were lost to attrition was statistically equivalent ( $M = 81.91, SD = 34.20$ ),  $t(155) = -0.779, p = .437$ . Similarly, among control group students, there was no significant difference in pretest fluency scores for children who remained in the study ( $M = 89.04, SD = 27.59$ ) and those who were lost to attrition ( $M = 79.2, SD = 17.25$ ),  $t(153) = -0.791$ .

### *Description of Experimental Conditions*

We implemented the READ 180 Enterprise Edition, Stage A (Elementary). There were 15 certified teachers in the READ 180 Enterprise program, who received both preservice and in-service professional development and met program certification requirements. Both the READ 180 Enterprise program and the district after-school program were implemented 4 days per week from October 2006 to April 2007 for approximately 23 weeks. During the first of 2 hours, children received a snack and homework assistance. During the second hour, children participated in either the READ 180 Enterprise program or the district after-school session. We adapted the full 90-minute READ 180 Enterprise

model to fit the 60-minute block of time in the after-school program.

Table 3 outlines the activities of READ 180 Enterprise and the district after-school program. On rotating days, READ 180 Enterprise students participated in whole-group lessons, during which teachers provided explicit reading comprehension and vocabulary instruction and modeled fluent reading. Students also participated in two of three rotations per day: individualized computer-assisted reading instruction, independent and modeled reading practice with leveled text, and teacher-directed reading lessons tailored to the reading level of small groups. The READ 180 Enterprise after-school program was designed to conclude with a teacher-directed whole-group wrap-up lesson to review key objectives from the lesson. As noted earlier, most components of the READ 180 Enterprise intervention were implemented with moderate to high fidelity. There was also no significant difference in the mean posttest reading scores (adjusted for pretest fluency scores) of high- and low-fidelity classrooms.<sup>1</sup>

Unlike READ 180 Enterprise, the district after-school program did not provide whole-group instruction, individualized computer-assisted reading instruction, or independent and modeled reading practice with leveled text. However, it did involve some small-group, teacher-directed lessons. Teachers could develop their own activities or choose from 16 activities, including informal arts projects, math games, and commercially developed program materials from InstaCamp themed kits (i.e., astronomy, space exploration, history, geography) and a math and literacy curriculum (KidzLit and KidzMath). A primary objective of this program was to encourage high attendance with engaging activities. The KidzLit research base is limited to one single-group pretest-posttest evaluation (Developmental Studies Center, 2003).

### *Empirical Strategy*

*ITT estimates.* To address our first research question, we used ordinary least squares (OLS) regression analysis to generate ITT estimates, which measure the impact of being offered the READ 180 Enterprise intervention. The empirical strategy is to identify the causal effect of



TABLE 3  
*Curriculum and Instructional Activities in 60-Minute After-School Programs*

Activity	READ 180 Enterprise	District After-School Program
Whole-group teacher-directed instruction	Teacher <ul style="list-style-type: none"> <li>• Provides explicit and systematic vocabulary and reading comprehension instruction</li> <li>• Models fluent reading</li> <li>• Uses nonfiction passages to activate prior knowledge and provide background information</li> </ul>	None
Individualized computer-assisted reading activities	Computer-assisted reading activities around content-area topics (i.e., people and cultures, science and math, history and geography) <ul style="list-style-type: none"> <li>• Reading Zone activates prior knowledge and provides embedded phonics instruction</li> <li>• Word Zone offers practice with fluent word reading</li> <li>• Spelling Zone provides support in spelling target words</li> <li>• Success Zone requires children to answer comprehension and make recording of their oral reading</li> </ul>	None
Independent/modeled reading of leveled books	Children read high-interest paperback books that match their Lexile levels and read along with audio books	None
Teacher-directed lessons tailored to reading level of small groups of students	Teachers provide differentiated vocabulary, reading comprehension, word reading, and fluency instruction in groups of no more than five students	Optional small-group teacher-directed lessons using 16 different activities, including <ul style="list-style-type: none"> <li>• InstaCamp: themed activity kits focused on history, geography, and space exploration</li> <li>• KidzMath: math practice, math games</li> <li>• KidzLit: trade books with teacher guides to develop vocabulary, discussion skills, and cultural awareness</li> </ul>
Teacher-directed whole-group wrap-up	Review of lesson	None

being randomly assigned to the READ 180 Enterprise after-school program regardless of the number of days students attended the program from October to April. Thus, the ITT estimates offer an unbiased estimate of the treatment effect if the READ 180 Enterprise program were implemented in an after-school setting and children were offered the opportunity to participate in the program. Using OLS regression, we

used the following model to generate an unbiased ITT estimate of READ 180 Enterprise:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 RB_i + \varepsilon_i \quad (1)$$

where  $Y_i$  represents the test score outcome for student  $i$ ,  $T_i$  represents whether the student was randomly assigned to READ 180 Enterprise or

the district after-school program,  $\mathbf{X}_i$  represents a vector of student characteristics including the pretest fluency score,  $RB_i$  represents the school by grade randomization block, and  $\varepsilon_i$  represents the error term. The coefficient  $\beta_1$  is the estimated difference in posttest scores between treatment and control students and represents the ITT estimate on each outcome measure.<sup>2</sup> We used OLS (Equation 2) to estimate the coefficients  $\beta_5$  and  $\beta_6$ , which capture the interaction between the treatment and grade level and indicates whether ITT estimates differ by grade level:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 \mathbf{X}_i + \beta_3 G4_i + \beta_4 G5_i + \beta_5 T_i \times G4_i + \beta_6 T_i \times G5_i + \beta_7 RB_i + \varepsilon_i \quad (2)$$

*TOT estimates.* Although the OLS models yield unbiased ITT estimates, they do not indicate whether treatment effects are larger for children who actively attended the READ 180 Enterprise after-school program from October to April. Because children with high attendance rates are likely to differ from children with low attendance rates on a number of observed and unobserved characteristics, the least squares estimate of READ 180 attendance is likely to be correlated with the error term in the test score equation. Therefore, to obtain the TOT estimate, we used initial random assignment status as an instrument for the total number of days that a student actually attended the READ 180 Enterprise after-school program. Initial random assignment is likely to be a valid instrumental variable if it predicts student attendance in READ 180 Enterprise, is uncorrelated with the residuals in the second-stage test score equation, and influences test scores exclusively through a student’s participation in READ 180 Enterprise (Angrist et al., 1996).

We used instrumental variables analysis in two stages. The fully specified first-stage model takes the form

$$Z_i = \pi_0 + \pi_1 \mathbf{X}_i + \pi_2 T_i + \pi_3 RB_i + \delta_i \quad (3)$$

where  $T_i$  represents initial random assignment status and serves as the instrument for READ 180 Enterprise attendance,  $\mathbf{X}_i$  is a vector of student background variables,  $RB_i$  denotes the school by grade randomization block, and  $\delta_i$  is

the error term that allows for the correlation of residuals among students in the same classroom. Using Equation 3, we estimated  $Z'_i$ , which is the predicted value in READ 180 Enterprise attendance on the basis of initial random assignment status. The second stage model is written as

$$Y_i = \beta_0 + \beta_1 \mathbf{X}_i + \beta_2 Z'_i + \beta_3 RB_i + \varepsilon_i \quad (4)$$

where the posttest reading score is predicted by  $Z'_i$  and the same independent variables that were included in the first-stage model. In the second-stage model, the coefficient  $Z'_i$  yields a TOT estimate that indicates the impact of attending the READ 180 Enterprise after-school program on the posttest reading outcomes. However, unbiased estimates of the TOT estimate rests on several assumptions. First, the instrumental variable must be random. This first assumption is likely to be met, because students were randomly assigned to conditions, which created two groups that did not differ on observable baseline characteristics (Table 2). Second, an instrumental variable must be correlated with the endogenous participation variable, that is, attendance in the READ 180 Enterprise after-school program. To test this assumption, we report  $F$  statistics from the first-stage model correlating the instrumental variable (i.e., random assignment to treatment) and the READ 180 Enterprise attendance variable. Some analysts have suggested that the  $F$  statistic from the first-stage model should, at minimum, exceed 10 to avoid bias associated with weak instruments (Gennetian, Morris, Bos, & Bloom, 2005; Stock, Wright, & Yogo, 2002). Third, an instrumental variable must be uncorrelated with the residuals,  $\varepsilon_i$ , in the second-stage model. The critical assumption here is that the instrumental variable must influence posttest reading scores exclusively and solely through students’ attendance in the READ 180 Enterprise intervention. Thus, the instrumental variable,  $T_i$ , is included as an independent variable in the first-stage model (Equation 3) but is excluded in the second-stage test score model (Equation 4). We used the “ivregress 2sls” routine in Stata Version 11.1 to estimate the TOT parameter estimates and standard errors in the second-stage model.<sup>3</sup>

## Results

Table 4 provides descriptive statistics on the posttest means and standard deviations on each respective outcome. On the SAT 10, the average national percentile rank was near the 46th percentile in vocabulary, comprehension, and spelling for READ 180 Enterprise students. The column denoting raw score differences for each outcome suggests that READ 180 Enterprise had higher attendance rates and higher scores on the posttest measure of fluency and spelling. Figure 2 displays month-specific attendance rates (i.e., the percentage of total days that students attended READ 180 Enterprise and the district after-school program), showing that attendance rates declined during the school year for students in both conditions.

*What Is the ITT Estimate of READ 180 Enterprise in a Voluntary After-School Program Designed to Improve the Reading Achievement of Low-Performing Children in the Upper Elementary Grades? Does the Impact of Offering the READ 180 Enterprise After-School Program Differ by Grade Level?*

Table 5 presents ITT estimates of READ 180 Enterprise on each posttest outcome. Model 1 includes the pretest fluency score as a covariate, and Model 2 includes both the pretest fluency score and the student demographic variables. Models 1 and 2 also include fixed effects for the randomization block from which children were randomly assigned to READ 180 Enterprise or control groups. The statistically significant impact estimator suggests that READ 180 Enterprise students outscored district after-school students by 8.43 scaled score points on reading vocabulary and 9.66 points on reading comprehension. Using the pooled posttest standard deviation for each outcome measure (see Table 4), we created effect size estimates by dividing the impact estimator by the unadjusted pooled within-group standard deviation reported in Table 4, yielding positive effect sizes in reading vocabulary ( $d = .23$ ) and reading comprehension ( $d = .32$ ). Finally, there was no statistically significant impact on spelling or oral reading fluency.

Table 6 reports results of regression analyses that include treatment-by-grade interaction terms. Model 2 includes the Grade 4-by-READ 180 Enterprise interaction term and the Grade 5-by-READ 180 Enterprise interaction term, which test the hypothesis that the ITT estimate of READ 180 Enterprise differed by grade level. The grade by treatment interactions revealed no statistically significant interactions and no consistent pattern of results across each of the respective reading outcomes. The most consistent finding reported in Table 6 suggests that READ 180 Enterprise had a positive and statistically significant impact on reading vocabulary and comprehension and that the impact did not differ across grade level. In addition, none of the models revealed statistically significant impacts on spelling and oral reading fluency. Taken together, the results from the ITT analyses in Tables 5 and 6 revealed a main effect of READ 180 Enterprise on reading vocabulary and comprehension scores and no interactions on the basis of student grade level.<sup>4</sup>

*What Is the TOT Estimate of READ 180 Enterprise in a Voluntary After-School Program Designed to Improve the Reading Achievement of Low-Performing Children in the Upper Elementary Grades?*

Table 7 displays TOT estimates of READ 180 Enterprise on each posttest reading measure. We used two-stage least squares to generate the parameter estimates and standard errors for the impact of attending READ 180 Enterprise on posttest reading scores. In the first-stage model (Equation 4), there was a strong relationship between the instrumental variable and the READ 180 Enterprise attendance variable, and the  $F$  statistic of 92.66 exceeded the cutoff of 10 used to identify weak instruments (Stock et al., 2002).

Estimation of the TOT estimates was generated by the second-stage model and yielded two findings. First, similar to the ITT estimates, there was a positive impact of attending READ 180 Enterprise on both vocabulary and comprehension posttest scores. The TOT estimates indicate that students who participated more actively in the READ 180 Enterprise intervention enjoyed

TABLE 4

*Posttest Scores by Condition for Full Sample and Grade-Specific Subsamples*

Variable	READ 180 Enterprise					District After-School Program				
	<i>n</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum	<i>n</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum
Full sample										
Attendance (total days)	155	65.10	22.33	0	84	157	58.09	27.31	1	84
DORF	151	106.42	27.01	36	181	146	103.73	24.48	34	175
SAT 10 scaled scores										
Spelling	149	631.75	31.28	540	715	146	625.88	37.85	476	753
Vocabulary	150	639.77	35.74	540	739	146	630.68	36.18	560	722
Comprehension	150	635.92	32.34	562	728	146	625.75	28.17	541	684
SAT 10 national percentile ranks										
Spelling	149	46.21	23.89	3	95	146	42.57	25.94	1	99
Vocabulary	150	44.95	22.80	2	97	146	38.47	24.29	3	96
Comprehension	150	45.83	24.09	2	98	146	38.31	22.33	1	90
Grade 4										
Attendance (total days)	55	67.69	19.39	0	84	57	66.42	24.15	2	84
DORF	54	104.98	25.51	59	163	55	101.13	25.70	45	175
SAT 10 scaled scores										
Spelling	52	619.48	32.59	540	703	55	613.45	42.85	476	753
Vocabulary	53	620.15	31.20	540	698	55	621.24	38.14	560	722
Comprehension	53	620.70	28.09	562	665	55	623.20	28.02	545	684
SAT 10 national percentile ranks										
Spelling	52	46.75	24.20	3	95	55	43.56	27.33	1	99
Vocabulary	53	42.30	19.23	4	89	55	42.60	22.91	8	96
Comprehension	53	42.28	22.03	5	78	55	44.42	21.28	2	90
Grade 5										
Attendance (total days)	70	62.31	24.77	0	84	68	56.38	27.58	3	84
DORF	68	109.75	27.55	36	181	64	108.67	20.40	58	147
SAT 10 scaled scores										
Spelling	68	636.59	29.63	577	715	64	634.14	35.61	554	741
Vocabulary	68	649.93	34.57	565	739	64	634.91	36.47	565	715
Comprehension	68	644.74	29.99	593	728	64	627.20	29.86	541	681
SAT 10 national percentile ranks										
Spelling	68	48.10	24.55	6	95	64	46.06	26.36	2	98
Vocabulary	68	50.07	24.41	3	97	64	39.39	26.19	3	93
Comprehension	68	51.15	24.10	10	98	64	38.09	23.20	1	82
Grade 6										
Attendance (total days)	30	66.87	21.30	6	84	32	46.88	28.15	1	84
DORF	29	101.31	28.34	52	150	27	97.33	29.08	34	160
SAT 10 scaled scores										
Spelling	29	642.38	26.13	594	710	27	631.59	23.94	587	688
Vocabulary	29	651.83	31.62	582	707	27	639.89	27.12	591	690
Comprehension	29	643.07	35.60	570	711	27	627.52	24.77	570	680
SAT 10 national percentile ranks										
Spelling	29	40.83	21.66	7	91	27	32.26	19.39	5	80
Vocabulary	29	37.79	22.85	2	81	27	27.89	19.60	3	69
Comprehension	29	39.83	25.84	2	91	27	26.37	17.60	2	71

Note. DORF = Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency; SAT = Stanford Achievement Test.

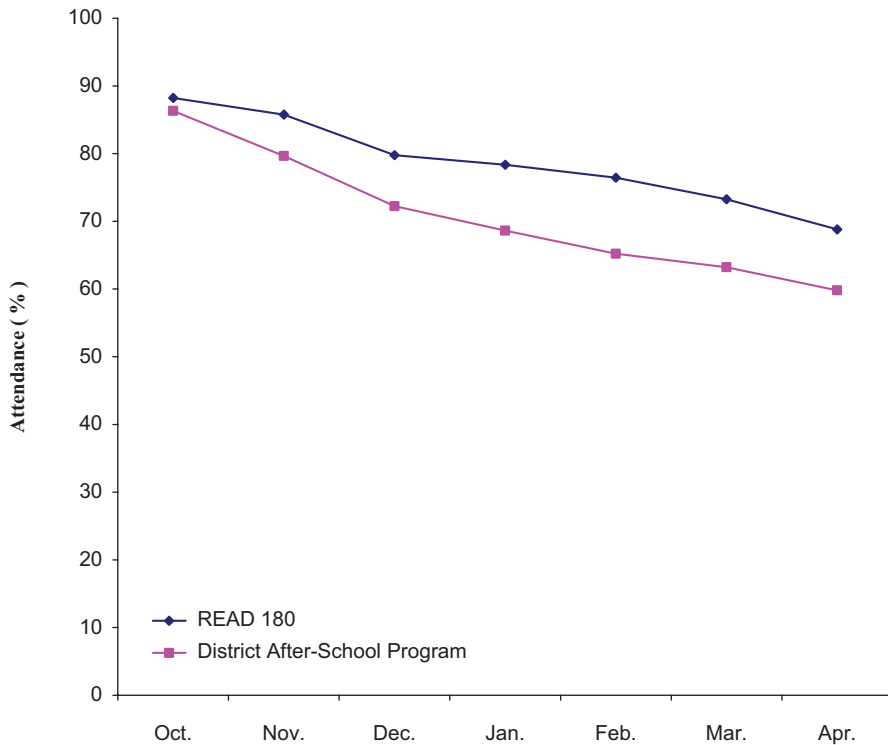


FIGURE 2. Percentage of attendance for all students enrolled in READ 180 Enterprise and district after-school program, by month.

TABLE 5  
Ordinary Least Squares Intention-to-Treat Estimates on Literacy Outcomes

Variable	Vocabulary		Comprehension		Spelling		Oral Reading Fluency	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
READ 180 Enterprise	8.406** (3.422)	8.433** (3.441)	9.426*** (2.987)	9.656† (2.946)	5.338 (3.433)	4.935 (3.424)	2.655* (1.485)	2.535* (1.491)
Pretest fluency score	0.734† (0.074)	0.739† (0.073)	0.602† (0.064)	0.578† (0.064)	0.687† (0.079)	0.672† (0.080)	0.883† (0.035)	0.878† (0.035)
Black student		-1.952 (3.916)		-2.175 (3.373)		1.245 (3.867)		-0.471 (1.678)
Latino student		-11.247** (5.606)		-4.233 (5.220)		1.026 (5.356)		0.253 (2.716)
Eligible for free lunch		0.500 (3.959)		-6.697* (3.437)		-5.236 (3.382)		-1.007 (1.684)
Female		-3.512 (3.710)		8.836*** (3.055)		-4.037 (3.703)		-2.647* (1.558)
Constant	566.541† (7.757)	568.953† (8.685)	569.667† (6.880)	573.246† (7.192)	557.730† (7.205)	562.601† (7.590)	31.186† (3.074)	33.257† (3.542)
<i>n</i>	296	296	296	296	295	295	297	297
<i>R</i> <sup>2</sup>	.368	.378	.332	.361	.323	.332	.764	.767

Note. Robust standard errors are in parentheses. All specifications include the student's school by grade randomization block. \**p* < .10. \*\**p* < .05. \*\*\**p* < .01. †*p* < .001.

TABLE 6

*Ordinary Least Squares Intention-to-Treat Estimates of Treatment-by-Grade Level Interactions on Literacy Outcomes*

Variable	Vocabulary		Comprehension		Spelling		Oral Reading Fluency	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
READ 180 Enterprise	8.433** (3.441)	8.603 (7.298)	9.656 <sup>†</sup> (2.946)	12.316* (7.448)	4.935 (3.424)	6.672 (6.019)	2.535* (1.491)	0.558 (3.441)
Pretest fluency score	0.739 <sup>†</sup> (0.073)	0.737 <sup>†</sup> (0.072)	.578 <sup>†</sup> (0.064)	0.576 <sup>†</sup> (0.064)	.672 <sup>†</sup> (.080)	.672 <sup>†</sup> (.080)	0.878 <sup>†</sup> (0.035)	0.879 <sup>†</sup> (0.035)
Black student	-1.952 (3.916)	-2.879 (3.939)	-2.175 (3.373)	-3.128 (3.369)	1.245 (3.867)	1.44 (3.831)	-0.471 (1.678)	-0.374 (1.719)
Hispanic student	-11.247** (5.606)	-11.115* (5.638)	-4.233 (5.220)	-4.163 (5.199)	1.026 (5.356)	.950 (5.390)	0.253 (2.716)	0.27 (2.713)
Eligible for free lunch	0.5 (3.959)	0.857 (3.952)	-6.697* (3.437)	-6.225* (3.417)	-5.236 (3.382)	-5.231 (3.452)	-1.007 (1.684)	-1.127 (1.694)
Female	-3.512 (3.710)	-4.794 (3.684)	8.836*** (3.055)	7.537** (3.043)	-4.037 (3.703)	-3.754 (3.762)	-2.647* (1.558)	-2.536 (1.587)
Grade 4	-16.814** (8.074)	-12.88 (10.150)	-15.166** (7.034)	-9.400 (8.626)	-6.962 (10.219)	-6.583 (11.148)	26.038 <sup>†</sup> (6.079)	24.394 <sup>†</sup> (6.381)
Grade 5	3.945 (9.258)	-0.283 (10.650)	-2.027 (7.455)	-4.804 (9.153)	9.090 (10.478)	11.163 (11.509)	42.423 <sup>†</sup> (6.257)	41.725 <sup>†</sup> (6.488)
READ 180 × Grade 4		-9.689 (9.142)		-13.196 (8.907)		-309 (8.947)		3.526 (4.159)
READ 180 × Grade 5		7.522 (9.079)		4.822 (8.576)		-3.613 (7.841)		1.528 (4.188)
Constant	585.767 <sup>†</sup> (10.063)	587.180 <sup>†</sup> (11.109)	588.412 <sup>†</sup> (9.586)	588.477 <sup>†</sup> (10.184)	569.563 <sup>†</sup> (12.908)	568.303 <sup>†</sup> (13.319)	7.219 (7.469)	8.051 (7.592)
<i>n</i>	296	296	296	296	295	295	297	297
<i>R</i> <sup>2</sup>	.378	.389	.361	.378	.332	.332	.767	.767

Note. Robust standard errors are in parentheses. All specifications include the student's school by grade randomization block.

TABLE 7

*Instrumental Variables Estimates of Attending READ 180 Enterprise on Literacy Outcomes*

Variable	Vocabulary	Comprehension	Spelling	Oral Reading Fluency
READ 180 Enterprise attendance	10.774** (4.280)	12.337*** (3.658)	6.301 (4.245)	3.233* (1.856)
Pretest fluency score	0.729 <sup>†</sup> (0.072)	0.566 <sup>†</sup> (0.063)	0.666 <sup>†</sup> (0.078)	0.875 <sup>†</sup> (0.034)
Black student	-2.189 (3.825)	-2.446 (3.286)	1.104 (3.755)	-0.541 (1.637)
Hispanic student	-11.067** (5.496)	-4.028 (5.080)	1.132 (5.161)	0.289 (2.646)
Eligible for free lunch	0.628 (3.866)	-6.551* (3.359)	-5.163 (3.289)	-0.970 (1.651)
Female	-3.406 (3.625)	8.957*** (2.976)	-3.977 (3.602)	-2.618* (1.523)
Constant	547.981 <sup>†</sup> (13.396)	565.386 <sup>†</sup> (12.549)	570.387 <sup>†</sup> (12.382)	7.623 (7.307)
<i>n</i>	296	296	295	297
<i>R</i> <sup>2</sup>	.375	.359	.331	.764

Note. Robust standard errors are in parentheses. All specifications include the student's grade level and school by grade randomization block.

\**p* < .10. \*\**p* < .05. \*\*\**p* < .01. <sup>†</sup>*p* < .001.

larger gains in reading vocabulary and comprehension than control students. Second, there was no evidence that attending READ 180 Enterprise was related to improvement in spelling and oral reading fluency, which is consistent with the ITT estimates reported earlier in Tables 4 and 5.

### **Discussion**

This study presents findings from a randomized controlled trial of READ 180 Enterprise in a voluntary after-school program designed to improve reading ability for mostly low-performing readers. The results of our randomized controlled trial suggest that implementation of the READ 180 Enterprise intervention in an after-school setting had a positive impact on reading comprehension and vocabulary scores on the SAT 10. The ITT estimates provide credible causal estimates of the impact of READ 180 Enterprise on student outcomes if the intervention was offered to a sample in which 95% of the children score below proficient on the English language arts MCAS. The magnitudes of the effect sizes of .23 in vocabulary and .32 in reading comprehension were in line with the mean effect size of .24 on reading comprehension tests from a recent review of eight quasi-experimental studies of READ 180 (Slavin et al., 2008). Furthermore, the estimates from the TOT analyses replicated the results from the ITT analyses. READ 180 Enterprise students who actively attended their after-school program performed better on reading vocabulary and comprehension. In sum, the key findings are robust across a number of model specifications and provide converging evidence that READ 180 Enterprise improved reading vocabulary and comprehension scores.

The results are also consistent with findings from an experimental evaluation of READ 180 (Lang et al., 2008, 2009) but differ from those from our previous study in which we found no effects in reading comprehension or vocabulary (Kim et al., 2010). The prior achievement of the student samples in these studies may help account for the similarities and differences in findings. The effect size of .32 in reading comprehension in this study mirrors Lang et al.'s reported effect size of .30 on the FCAT for students whose average score on the norm-referenced

SAT 9 comprehension test was at the 44th percentile. Similar to the Florida sample, our current sample included children who scored at the 46th percentile on the SAT 10 reading comprehension posttest. In contrast, students in our first study were lower performing, scoring at approximately the 25th percentile in comprehension. Just as our previous study found no effect on reading comprehension for this group, Lang et al. (2009) also found no effect of READ 180 for the sample of high-risk students scoring at the 24th percentile.

However, a difference in prior achievement is not the only factor that distinguishes our current and previous studies. The different versions of READ 180 used in our two evaluations—READ 180 Version 1.6 in the previous study versus READ 180 Enterprise Edition in the current study—may also help account for the varying effects. Although both studies implemented a 60-minute version of READ 180 in an after-school program, only the current study included regular teacher-directed whole-group instruction and whole-group wrap-up. These teacher-directed whole-group activities offered students systematic and explicit instruction in vocabulary, as well as opportunities to deepen content area knowledge. The Simple View of Reading (SVR) (Gough & Tunmer, 1986) helps interpret the impact of this additional instructional emphasis on language comprehension. According to the SVR, reading comprehension is the product of word recognition (i.e., decoding, fluency) and language comprehension (i.e., vocabulary). Because the teacher-directed whole-group instruction provided students with an opportunity to build vocabulary, it is likely that these activities had a positive impact on reading comprehension.

Taken together, our two studies and Lang et al.'s (2008, 2009) evaluation may shed new light on the READ 180 Enterprise logic model (Figure 1). According to Scholastic, the program is appropriate for students in the lower 25th percentile and most effective when both whole-group instruction and the three rotations are implemented. Consistent with this logic, our evidence suggests that a combination of whole-group instruction and rotations yields positive effects. However, findings from the three experiments suggest that significant gains on standardized

assessment may be observed for moderate risk, as opposed to the most struggling readers. Thus, the most recent experimental findings suggest that READ 180 can improve student outcomes if (a) it targets moderate risk students scoring near the 40th to 45th percentile and (b) it implements both teacher-directed whole-group instruction and the three small group rotations.

Despite the positive impact on vocabulary and reading comprehension, the ITT estimates also suggest that READ 180 Enterprise had no significant impact on spelling or oral reading fluency. What explains these null findings? We hypothesize that the amount of time dedicated to spelling may have been insufficient to affect scores on a standardized assessment. READ 180 Enterprise targets spelling development almost exclusively through individualized computer-assisted reading instruction, which accounts for only a small fraction of the program. More time on spelling tasks may have been required to affect the spelling performance of students scoring at the 44th percentile on a standardized measure. Unlike spelling, fluency is a primary focus of this intervention (Lang et al., 2009). However, students may have spent too little time on the fluency activities most likely to affect change. The 60-minute version of READ 180 Enterprise in this study offered multiple opportunities for modeled reading through teacher instruction, computer-based activities, and audio books but fewer opportunities for students to read aloud themselves and receive guided feedback, two factors that are associated with improved fluency (National Institute of Child Health and Human Development, 2000). Students may have needed more opportunities to read aloud and receive feedback than the intervention offered.

In addition to the implications of this study that are specific to READ 180, this randomized experiment also has policy implications for evaluations of after-school programs. Our findings suggest that literacy intensive supplementary education programs are more likely to improve reading comprehension outcomes than those with a less specific focus on literacy instruction. The fact that the effect size of .32 in reading comprehension is nearly three times the effect ( $d = .13$ ) reported in a meta-analysis of the impact

of after-school programs on reading achievement (Lauer et al., 2006) may be explained by the intensive literacy focus of READ 180 Enterprise. Our findings also suggest that evaluation of mixed-method literacy interventions such as READ 180 require measurement of multiple outcomes that the intervention is designed to improve.

### *Limitations and Future Research*

Three limitations of this study should be addressed in future work. First, replication efforts should test the READ 180 Enterprise logic model (Figure 1). Although Scholastic recommends that schools target students in the bottom quartile, results from our previous and current study, together with the evaluation by Lang et al. (2008, 2009), are suggestive that the program in fact may be better targeted toward moderate-risk students reading just below national norms on standardized assessments. Further study is needed to determine whether READ 180 Enterprise's logic model in Figure 1 might be more accurately represented by changing the targeted sample from "high-risk readers" to "moderate-risk readers." Specifically, given that our current and previous studies differed both in the students' levels of prior achievement and the addition of whole-group instruction, future research should examine whether one or both of these differences are responsible for influencing change.

Second, replication is needed to test the external validity of the findings from this study and to identify program components that are likely to improve student achievement. Previous evaluations of after-school programs serving elementary grade students have shown mixed effects on standardized test scores (Kane, 2004; Lauer et al., 2006; Zief et al., 2006). In the National Evaluation of the 21st Century Community Learning Centers Program, James-Burdumy et al. (2007) conducted both ITT and TOT analyses and found no significant effect on the SAT 9 reading scores. On the basis of these findings, James-Burdumy et al. hypothesized that a focus on "strengthening lagging competencies during afterschool time could possibly improve academic outcomes more than programs that provide many types of activities that may be appealing to a range of students but do not



focus on particular skill areas” (p. 314). Although after-school programs serve a number of social, personal, and academic goals, developmental theory underscores the importance of implementing programs with sequential learning activities that promote active, focused, and explicit learning opportunities among participating students (Durlak & Weissberg, 2007; Granger, 2008; Granger, Durlak, Yohalem, & Reisner, 2007). The READ 180 Enterprise program seems to incorporate many of these principles in an after-school setting. It should also be emphasized that the goals of the district after-school program (i.e., the counterfactual condition) in this study were similar to those in the 21st Century Community Learning Centers, which offer students a mix of recreational, social, and academic activities. Nonetheless, in the absence of a large-scale evaluation, it is unclear whether READ 180 Enterprise or other after-school programs with a focused academic program could improve achievement at scale in a large and diverse sample of school districts.

Finally, there is a need to conduct longitudinal studies that explore whether children’s attendance in after-school programs and reading achievement change over time. Our study indicated that attendance rates were higher for Grade 4 children than Grade 5 and Grade 6 children. However, our study is based on three grade specific cohorts and does not permit individual growth modeling of key motivational and cognitive outcomes. Ultimately, improving attendance in a high-quality, structured literacy program such as READ 180 Enterprise may enhance students’ ability to read for understanding in the upper elementary and middle grades.

### Notes

1. We used a multilevel model to examine whether average fidelity scores predicted posttest reading scores controlling for pretest fluency scores. The Level 2 (classroom-level) model included the posttest reading score, a fidelity dummy variable (1 = high-fidelity classroom, 0 = low-fidelity classroom), and pretest fluency scores. The coefficient for the fidelity dummy variable was insignificant in each of the four models. Although the coefficient for the dummy variable on the fidelity score was not statistically significant in each of the four models, there was suggestive evidence that classrooms with fidelity scores above the median scored higher, on average, than classrooms with low fidelity

scores on posttest measures of reading vocabulary ( $d = .10$ ), comprehension ( $d = .26$ ), spelling ( $d = .18$ ), and fluency ( $d = .45$ ). These  $d$  indexes are based on covariate-adjusted (i.e., fluency pretest scores) posttest differences on each of the four reading outcome measures.

2. We used robust standard errors (White, 1980) to account for the correlation of residuals among students in the same school by grade randomization blocks.

3. As noted by Angrist and Pischke (2009), the correct residual variance estimator “uses the original endogenous regressor to construct residuals and not the first-stage fitted values” (p. 140). These adjustments are made automatically in the Stata “ivregress 2sls” routine.

4. In addition to testing for interactions between the treatment and grade level, we also examined whether the impact of offering the READ 180 Enterprise program differed for students who scored above and below the grade-specific median on the pretest fluency measure, our only available pretest measure. We thank an anonymous reviewer for suggesting this analysis. To conduct this analysis, we included a two-way interaction between experimental condition (1 = READ 180 Enterprise, 0 = district after-school) and a binary variable denoting the performance of low-performing students on the grade-specific fluency pretest score (1 = below the grade-specific median, 0 = above the grade-specific median). None of the interaction terms between pretest reading fluency score and treatment was statistically significant in the analyses of the four literacy outcome measures. We interpret this finding with caution, however. The pretest DORF is not a measure of general reading ability or reading comprehension. As a result, a student who scores low on the DORF is not necessarily a struggling reader. Thus, we cannot conclude that the effect of READ 180 Enterprise does not differ for good and poor readers but only that there is no evidence that READ 180 Enterprise had differential effects based on students’ initial fluency scores. To determine whether READ 180 Enterprise has a differential effect for students of varying reading abilities, researchers should use a pretest measure of general reading ability or reading comprehension as a component of an interaction term.

### References

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(234), 444–455.
- Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist’s companion*. Princeton, NJ: Princeton University Press.
- Borenstein, M., Hedges, L. V., Higgins, J.P.T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*:

- Statistics in practice*. Chichester, UK: Wiley.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230.
- Buly, M. R., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, 24(3), 219–239.
- Butcher, K. R., & Kintsch, W. (2003). Text comprehension and discourse processing. In A. F. Healy & R. W. Proctor (Vol. Eds.) & I. B. Weiner (Ed.-in-Chief), *Handbook of psychology, Volume 4: Experimental psychology* (pp. 575–595). New York: John Wiley.
- Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York: Carnegie Corporation of New York.
- Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.
- Chicago Public Schools. (2009). CPS expands READ 180 program in 80 elementary schools. Retrieved January 17, 2010, from [http://www.cps.edu/News/Press\\_releases/2009/Pages/11\\_30\\_2009\\_PR1.aspx](http://www.cps.edu/News/Press_releases/2009/Pages/11_30_2009_PR1.aspx)
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, 22(1–2), 8–15.
- Deschler, D. D., Palinscar, A. S., Biancarosa, G., & Nair, M. (2007). *Informed choices for struggling adolescent readers: A research-based guide to instructional programs and practices*. Newark, DE: International Reading Association.
- Developmental Studies Center. (2003). *Summary of after-school KidzLit evaluation*. Retrieved January 17, 2010, from <http://in.devstu.org/page/research-results-1>
- Durlak, J. A., & Weissberg, R. P. (2007). *The impact of after-school programs that promote personal and social skills*. Retrieved January 2, 2010, from <http://www.casel.org/downloads/ASP-Full.pdf>
- Gennetian, L. A., Morris, P. M., 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: Russell Sage.
- Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills*. Eugene, OR: Institute for the Development of Educational Achievement.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, 7(1), 6–10.
- Granger, R. (2008). After-school programs and academics: Implications for policy, practice, and research. *Social Policy Report*, 22, 1–19.
- Granger, R., Durlak, J. A., Yohalem, N., & Reisner, E. (2007, April). *Improving after-school program quality*. New York: William T. Grant Foundation.
- Guthrie, J. T. (2003). Concept oriented reading instruction. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 115–140). New York: Guilford.
- Hartry, A., Fitzgerald, R., & Porter, K. (2008). Implementing a structured reading program in an afterschool setting: Problems and potential solutions. *Harvard Educational Review*, 78(1), 181–210.
- Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading and Writing Quarterly*, 20(2), 123–144.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2(2), 127–160.
- James-Burdumy, S., Dynarski, M., & Deke, J. (2007). When elementary schools stay open late: Results from the National Evaluation of the 21st Century Community Learning Centers Program. *Educational Evaluation and Policy Analysis*, 29(4), 296–318.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80(4), 437–447.
- Kane, T. J. (2004, January). *The impact of after-school programs: Interpreting the results of four recent evaluations* (W. T. Grant Foundation Working Paper). Retrieved February 2, 2010, from [http://www.wtgrantfoundation.org/usr\\_doc/After-school\\_paper.pdf](http://www.wtgrantfoundation.org/usr_doc/After-school_paper.pdf)
- Kieffer, M. J., & Lesaux, N. K. (2007). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *Reading Teacher*, 61(2), 134–144.
- Kim, J. S., Samson, J. F., Fitzgerald, R., & Hartry, A. (2010). A randomized experiment of a mixed-methods literacy intervention for struggling readers in grades 4–6: Effects on word reading efficiency, reading comprehension and vocabulary, and oral reading fluency. *Reading and Writing*, 23(9), 1109–1129.
- Lang, L. H., Torgesen, J. K., Petscher, Y., Vogel, W., Chanter, C., & Lefsky, E. (2008, March). *Exploring the relative effectiveness of reading interventions for high school students*. Paper presented at the annual research conference of the Society for Research on Educational Effectiveness, Crystal City, VA.
- Lang, L., Torgesen, J., Vogel, W., Chanter, C., Lefsky, E., & Petscher, Y. (2009). Exploring the relative effectiveness of reading interventions for high school students. *Journal of Research on Educational Effectiveness*, 2(2), 149–175.
- Lauer, P. A., Akiba, M. A., Wilkerson, S. B., Apthorp,

- H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*(2), 275–313.
- Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007* (NCES 2007-496). Washington, DC: National Center for Education Statistics.
- Loveless, T. (2007). *How well are American students learning?* Washington, DC: Brown Center on Educational Policy at Brookings.
- Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., . . . Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90*(3), 347–373.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. W. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Scholastic, Inc. (n.d.). *About READ 180*. Retrieved November 30, 2009, from <http://teacher.scholastic.com/products/read180/overview/faq.htm#instr02>
- Scholastic Research. (2008). *Desert Sands Unified School District, CA*. New York: Scholastic Inc.
- Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly, 43*(3), 290–322.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407.
- Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics, 20*(4), 518–529.
- Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (Rev. ed.). Alexandria, VA: ASCD.
- Torgesen, J. K. (2005). Recent discoveries on remedial interventions for children with dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 521–537). Oxford, UK: Blackwell.
- U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. (2007). *State and local implementation of the No Child Left Behind Act, Volume I—Title I school choice, supplemental educational services, and student achievement*. Washington, DC: Author.
- U.S. Department of Education, Office of the Undersecretary. (2003). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program* (First year findings). Retrieved November 9, 2006, from <http://www.ed.gov/pubs/21cent/firstyear/index.html>
- What Works Clearinghouse. (2009). *Intervention READ 180*. Washington, DC: U.S. Department of Education.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817–838.
- Zief, S. G., Lauer, S., & Maynard, R. A. (2006). *The impacts of after-school programs on student outcomes: A systematic review for the Campbell Collaboration*. Retrieved September 7, 2009, from [http://www.campbellcollaboration.org/reviews\\_education/index.php](http://www.campbellcollaboration.org/reviews_education/index.php)

## Authors

JAMES S. KIM is Assistant Professor, Harvard University Graduate School of Education, 14 Appian Way, Larsen 505, Cambridge, MA 02138; kimja@gse.harvard.edu. His areas of expertise include experimental evaluations of literacy interventions and the use of social science evidence in education policy.

LAUREN CAPOTOSTO is a doctoral student at the Harvard Graduate School of education. Her research focuses on adult and family literacy.

ARDICE HARTRY is a Research Specialist with the Lawrence Hall of Science Research Group at the University of California, Berkeley

ROBERT FITZGERALD is a senior research associate with MPR Associates, a public policy research and evaluation firm specializing in education. Currently, he is responsible for the analysis and validation tasks on a Gates Foundation funded three state pilot projects to develop high-quality, actionable, data-driven reports linking K–12 and postsecondary education data that can be used by schools, districts, and states to improve the