Afterschool Research Brief

The National Partnership for Quality Afterschool Learning Randomized Controlled Trial Studies of Promising Afterschool Programs: Summary of Findings

Michael Vaden-Kiernan, PhD, Debra Hughes Jones, PhD, and Zena Rudo, PhD, SEDL
Robert Fitzgerald, MA, and Ardice Hartry, PhD, MPR Associates, Inc.
Bette Chambers, PhD, and Dewi Smith, MA, Success for All Foundation
Patricia Muller, PhD, and Marcey A. Moss, PhD, The Center for Evaluation and Education Policy (CEEP) at Indiana University

Introduction

Recent evaluations and research syntheses of afterschool programs rated as high quality show they are associated with increases in student achievement and other positive socio-behavioral outcomes (Lauer et al., 2006; Vandell, Reisner, & Pierce, 2007). Those examinations provide a springboard for the next much-needed area of investigation—whether afterschool programs containing academic content can have positive impacts on student achievement—about which scant rigorous evidence exists. In an effort to answer questions about the impact of promising afterschool interventions containing academic content, the U.S. Department of Education funded three randomized controlled trial (RCT) studies to evaluate potential benefits on student achievement.

As part of SEDL's National Partnership for Quality Afterschool Learning, the award competition to fund three RCT projects was coordinated, along with a plan to facilitate the technical and analytic support of the research projects through the Afterschool Research Consortium (ARC). The ARC was composed of a subgroup of methodological and afterschool experts that had been brought together to review proposals, along with SEDL researchers, and key staff from each afterschool research project. The ARC was conceived as a "cutting-edge" opportunity to collaborate, rather than compete, in applying best research practices and address important challenges to support awardees' efforts over the 2-year funding period (2006–2008). The consortium provided an open forum for members to discuss challenges, solutions, and accomplishments in a supportive, collegial setting in order to advance the effective use of rigorous experimental research approaches in applied afterschool settings.

The RCT project staff submitted final reports for each of their study's findings in Fall 2008, and they are the source of the information synthesized and presented here. This afterschool research brief, the final in a three-part series,[1] presents an overview of the studies and a summary of implementation and impact findings across the 2 years of funding.

Project Background

In Summer 2006, SEDL funded three RCT efficacy trials of promising literacy interventions, implemented in afterschool settings, on elementary students' reading achievement. Promising afterschool interventions were defined as those that were fully developed, had already been implemented in an education setting, were replicable, and for which a strong case could be made that the study of such an intervention would have important implications for practice and policy. The interventions were to target elementary-

---

[1] *Afterschool Research Brief* no. 1 provides detailed information about the study criteria, the selection process, and a discussion of the state of the field of afterschool research using RCT designs. *Afterschool Research Brief* no. 2 presents the primary challenges undertaken by the studies, which concerned difficulties with the recruitment of sites and challenges with implementation of curricula that had been adapted to fit the afterschool setting.

aged students with a focus on improving student academic outcomes. These efficacy trials were to be conducted under ideal conditions, with support from the program developers, to support implementation of the program components as theorized to affect outcomes, and thereby to evaluate whether they indeed have any effect.

Brief Overview of the Programs

The interventions evaluated in these trials, Adventure Island, Voyager Passport, and READ 180, incorporate the five essential components of reading instruction: phonemic awareness, phonics, fluency, vocabulary, and comprehension. These reading component standards, as laid out in the National Reading Panel's 2000 report, *Teaching Children to Read,* its companion piece for teachers, *Put Reading First* (2001), and the 1998 report of the National Research Council, *Preventing Reading Difficulties in Young Children,* were derived from scientifically based research studies. The reading programs use multiple strategies, at levels matched to students' needs, for the delivery of instructional materials targeted to struggling readers. The curricula and supporting empirical evidence are summarized briefly in this section.

Success for All's Adventure Island

Success for All's (SFA) Adventure Island is an afterschool reading program based on the SFA reading model (Slavin & Madden, 2001) with a focus on the components that are identified as common deficiencies among struggling readers, such as phonics, fluency, and metacognitive comprehension strategies. It also highlights components of particular importance to English language learners, such as vocabulary building. Adventure Island is highly social, motivating, and engaging, and it is designed around a common theme that emphasizes adventures at sea, discovery, and treasures (Slavin, 1995). Students are assigned to four-member teams, with a structure for team recognition (e.g., certificates) and other rewards if all members do well on individual assessments, emphasizing cooperative learning methods. At the beginning of SEDL funding, Adventure Island was one of several curricula included in a national RCT-designed evaluation of afterschool programs (Black, Doolittle, Zhu, Unterman, Grossman, 2008); preliminary findings indicate no statistically significant impacts on reading performance for the Adventure Island reading program after the first

program year. Although the research on Adventure Island is limited, the SFA reading program has been evaluated extensively. Recent research involving a national randomized evaluation of SFA in grades K–2 found significant positive effects of the program across 35 high-poverty schools (Borman et al., 2005, 2007). Teacher professional development training for Adventure Island provides instruction in strategies such as the use of realia (i.e., use of concrete objects to support learning objectives), choral responding, repetition, elaboration, pantomime, total physical response, and other supportive practices for teaching English as a second language (August & Shanahan, 2006; Carlo et al., 2004; Calderón, 2001). Developers offer supports to teachers using the program: professional development, follow-up technical assistance, in-class visits, mentoring by school facilitators or district coordinators, and a mid-year conference.

Voyager Expanded Learning's Voyager Passport

Voyager Expanded Learning's reading program, Voyager Passport, is designed to accelerate the reading performance of struggling readers to grade level by providing systematic lessons aimed at strengthening reading skills. Voyager Passport provides explicit and systematic instruction through the implementation of two components in every lesson; "Word Works" provides grade-appropriate instruction in phonemic awareness, letter-sound recognition, word reading, and sight words. The "Read to Understand" component gives struggling readers daily opportunities to apply newly learned skills with accessible and engaging text. Fluency practice is provided through the "Extra Fluency Readers." Students also work in small groups, increasing their vocabulary use through wide reading and interactions; semantic maps and other graphic organizers are also provided to help students connect concepts to words. Voyager Passport is based on the struggling reader intervention component in the Voyager Universal Literacy System, which is supported by evidence compiled from various studies using matched control designs (Frechtling, Silverstein & Zhang, 2003; Hecht, 2003; Hecht & Torgensen, 2002; Roberts & Allen, 2003). Voyager has provided extended day and summer intervention programs to more than 750,000 students nationwide; Voyager Passport has been implemented as a summer program in several pre-post test designs, with preliminary evidence of its success

with the academic growth of low performing students. The Voyager Passport system includes training and technical assistance for implementation and provides pre-launch planning, teacher training and professional development materials, consultations with principals and coaches, on-going telephone and e-mail support, and supplemental online support materials and product training.

Scholastic, Inc.'s READ 180

Scholastic, Inc.'s READ 180 program, based on the work of Dr. Ted Hasselbring at Vanderbilt University and more than a decade of research on literacy, is delivered through three different session types. The first is a 20-minute whole-class direct instruction session, after which three small-group rotations of 20 minutes each are begun for small-group direct instruction, independent and modeled reading, and use of READ 180 topic software at individual computer stations. The third session type is 10 minutes of whole-class direct instruction to conclude the class. READ 180 has been the subject of extensive research in the regular day school setting, including quasi-experimental, correlational, and descriptive studies. Based on its success in the regular day-school classroom, the program was adapted to the afterschool setting and tested in an RCT study to examine the impact of READ 180 in afterschool classrooms in Brockton, Massachusetts. Findings from the study, funded by the William T. Grant Foundation, revealed that READ 180 had an impact on the reading skills of the 150 students in the treatment group in three elementary schools in the study (Hartry, Fitzgerald, & Porter, 2008). READ 180 includes pre-service and in-service teacher professional development, audiobooks, paperbacks, and topic software for READ 180. Teachers received a full day of training prior to launch, follow-up training 6 weeks later (e.g., half-day training), a "cadre meeting" several times a year, which is facilitated by trainers to discuss problems and find solutions, and access to Scholastic Red, an online professional development program (Slavin, Groff, & Lake, 2008).

Methods

Research Questions

The ARC, involving SEDL researchers, awardees, and experts in the field,[2] contributed to the development of a set of experimental research questions and a secondary correlational research question focused on implementation. Generally, the research questions guiding the set of studies are as follows:

Experimental Questions

1. Does the reading intervention (treatment) improve students' reading skills more than typical afterschool activities?

2. Does the reading intervention (treatment) improve student outcomes related to academic achievement, such as afterschool attendance and attitudes towards reading?

3. Does the reading intervention (treatment) work equally well for different subgroups of students, including students who vary according to ethnicity/race, grade level, reading abilities, and gender?

Correlational Question

1. What is the relationship between fidelity of implementation of the reading intervention (treatment) and student outcomes?

Study Design

The studies were conducted using experimental design methodology that employed random assignment of students, within afterschool sites, to either treatment or control groups. The evaluations included descriptive implementation information about challenges associated with integrating structured academic content in typical afterschool programs and settings. Assessments were administered to students in the fall and spring of each year, and information on afterschool students, instructors, and classrooms were collected several times per year.

---

[2] Fred Doolittle, MDRC; Elizabeth Reisner, Policy Studies Associates, Inc.; and Peter Witt, Texas A&M University participated in the ARC as Technical Working Group members.

Samples and Assignment to Condition

Afterschool sites participating in the evaluation of the reading curricula varied in several important ways—the broader context from which the sites were drawn, the characteristics of the specific afterschools sites, and program administrators' background experience and relationship with the evaluation researchers. The program contexts included predominantly Hispanic, large, urban districts in a Southwestern state (SFA); a mostly rural sample of 21st Century Learning Community Center programs from across an entire state in the Midwest (CEEP); and a large suburban district on the fringe of a major Northeastern metropolitan area (MPR).

Each study design used the procedure of randomly assigning students to treatment or control classrooms within afterschool program sites. The random assignment procedure began by forming a pool of eligible elementary students who returned parent-signed agreements to participate in the study and who formally enrolled in each respective afterschool program. A few weeks before the study began, students' baseline scores on reading ability were assessed. Within each site, students were stratified by grade and gender (and, in the SFA study, by English Language Learner status), and were then numbered and assigned to condition using a random numbers table. After students were randomly placed into the treatment or control groups the instruction began.

Characteristics of Students in the Evaluation

The variation between study contexts is reflected in the student demographics of the samples used for the impact analyses. For example, the SFA study's sample was mostly Hispanic, the MPR sample was predominantly African American, and the CEEP sample was almost all White. The average percent of low-income students ranged from 52 to 91%, with the lowest percentage in the CEEP sample. The student

Table 1. Overview of Student Demographic Information for Each Study

	Average % Free- Reduced-Lunch	Average % Minority Student Enrollment	Average % Female	Number of Students and Sites
SFA Adventure Island	76	90	47	T: 242 C: 242 S: 5
CEEP Voyager Passport	52	7	50	T: 119 C: 133 S: 15
MPR READ 180 Year 1	68	72	54	T: 155 C: 157 S: 4
MPR READ 180 Year 2[3]	91	65	53	T: 152 C: 152 S: 4

NOTE: T: Treatment Group
C: Control Group
S: Sites

[3] MPR collected 2 years of data; SFA was unable to recruit enough sites to begin the study in the 1st year and so has 1 year of data in Year 2; CEEP's study was concluded at the end of the 1st year due to unresolved issues between the evaluation research team and developer.

demographic information for each study is summarized in Table 1.

Additionally, in the SFA Adventure Island study, 56% of the students in the sample were fluent English speakers and 44% were labeled "English Language Learners" (ELL). Among the ELL students, Spanish-speakers represented the largest language group (approximately 32%) with the remainder categorized as Vietnamese, Arabic, Tagalog, and Igbo speakers.

Implementation Measures

The project teams collected data from multiple sources using a variety of methods to assess study implementation and fidelity of program implementation. The implementation measures included observational rating scales that were used during site visits. Information about teachers' use of the instructional components of the programs was collected using these rating scales during the scheduled structured observations. Focus groups and interviews with principals, afterschool program directors, and instructors also were conducted during the site visits; these sessions provided descriptive information about teachers' experience with the intervention in their afterschool classrooms, information about the quality of the professional development they received and its relationship to implementation of the intervention; administrators reported on their general perceptions of how well program implementation went and any other special issues that arose during the study.

Student Outcome Measures

The outcomes used in the impact analysis for each study were measured at the individual student level. Two of the three studies used Harcourt's Stanford Achievement Test (SAT 10) for the primary outcome measure (total reading score; word skills, vocabulary, comprehension, spelling); the Adventure Island study used the Woodcock Johnson Test of Achievement, III and DIBELS oral fluency score as the principal outcome measures. All SAT 10 test scores are scaled scores therefore the scores can be compared across grades.

Analytic Approaches

The SFA Adventure Island and CEEP Voyager Passport studies used multivariate analysis of covariance (MANCOVA) for their main impact analysis. In order to answer the main research question of impact, SFA conducted separate MANCOVAs for the intent-to-treat (ITT) sample and the treatment-on-the-treated (TOT) samples. The PPVT and the WJ-III were used as covariates in the analyses. The CEEP Voyager Passport study also used MANCOVA models in ITT analyses to determine if students participating in the treatment group performed differently than students participating in regular afterschool activities, using two SAT 10 subscales, vocabulary and comprehension, as the outcome measures. Controls were included in the model to account for differences between the two groups; controls included prior reading ability (DIBELS "pre-test" score), a proxy for socio-economic status (i.e., free/reduced-price lunch eligibility), student grade level, gender, and special education status. In the MPR study, differences in assessment outcomes between the READ 180 and regular afterschool groups (ITT samples) were estimated using ordinary least squares regression for both years of data. The regression model estimated the outcomes after accounting for the baseline achievement test, measured at the beginning of the school year prior to random assignment, a dummy variable to indicate whether the student was assigned to the treatment or control group, and additional dummy variables for blocking factors used in random assignment.

Power Analyses

Power analyses of the analytic (ITT) samples indicate that all three studies had adequate power to detect effect sizes of +0.22 or smaller. Power estimates for the final samples were calculated for 80% power to detect the following minimum effects: the Adventure Island study—with a sample of 484 students—could detect an effect size of +0.15, using a covariate that explains 67% of posttest variation; the Voyager Passport study—with a sample of 252 students—could detect an effect size of +0.22, using a covariate that explains 61% of posttest variation; the READ 180 study—with samples of 312 (Year 1) and 304 (Year 2) students—could detect impacts at +.20, using a covariate that would explain 61% of posttest variation.

Implementation and Impact Findings

The key program implementation findings described here include how well the elements of the reading

programs were installed and a brief summary of the implementation challenges that were encountered across the studies. The impact findings for each study are then described and summarized.

Program Implementation

Findings for implementation of the SFA Adventure Island study were characterized by the acceleration of improvement in instructor skills over the school year. By the end of the study period, one study site had achieved levels ranging from routine or higher on implementation of program components. In that afterschool site, instructors were improving on delivering components of the material and also continued to refer to their training manual during the session for support. In the other Adventure Island afterschool sites, instructors were still learning strategies for delivery of the program materials. Descriptive analyses showed no significant relationship between the level of implementation of Adventure Island and student achievement. The findings also revealed the three most challenging program elements to implement: cooperative learning, facilitating student discussion at a challenging level, and supporting ELL students' learning. These data indicate that after 1 year of implementation the program was not being implemented fully with fidelity in all the afterschool sites.

In CEEP's Voyager Passport study, there was wide variation in implementation quality across sites and within a given site. From the outset of the study, difficulties arose with sites' unwillingness to devote the needed amount of days per week and time per day to the program. These challenges were never overcome. One of the most prevalent challenges to high-quality program implementation was inconsistency in the quality of instruction observed across instructors. Many instructors were unclear about the different program levels and their purpose, and therefore were more likely to use materials incorrectly, omit certain program modules, or have difficulty with following the teacher manual. Aside from the challenges associated with the quality of instruction, there were several challenges related to study implementation, including problems with final site recruitment and the associated substantial decrease in initial sample size, site attrition after the initiation of the study, and low student attendance.

In MPR's READ 180 study, modifications were made to the program between Year 1 and 2 of the study. The adaptations were in response to afterschool program and student activity scheduling requests, which initiated a change from 4 days per week, 60 minutes per day, to 2 days per week, 90 minutes per day. Afterschool, evaluation research, and developer staff were all informed of and agreed to the program changes. Analyses comparing 2nd-year results for one of the four schools that retained the 4 day per week model to the other three schools that adopted the 2 day per week model revealed no support for the possibility that the program modifications obscured positive impacts of READ 180. Instructors consistently endorsed as an improvement the program modification to 2 days per week, and generally gave high ratings to the program teacher training, supports, and general perceptions of the program.

Program Impact

Given the challenges to reaching high-quality implementation faced by the SFA and CEEP studies, it may not be a surprise that neither study found significant program impacts on student reading outcomes after 1 year. The only study that found significant impact effects, and only for the 1st year, was the READ 180 study. In Year 1, READ 180 had large and statistically significant impacts on SAT 10 vocabulary, comprehension, and total reading. On average, READ 180 students scored 8.5 points higher on the vocabulary, almost 10 points (9.50) higher on reading comprehension, and 15 points higher on total reading than control group students (total score limited to 5th and 6th graders). The comparable effect sizes (in standard deviation units) were, for vocabulary gains, almost one quarter of a standard deviation, comprehension gains were .31 standard deviations, and total reading more than one-half (.55) standard deviations. The overall impact of READ 180 on student reading outcomes is unclear given the findings from the 2 years of the study. The lack of any statistically significant effects for the READ 180 students in Year 2 was unexpected given the number of significant effects on reading outcomes from the impact analysis in the previous year.

The findings for the READ 180 study were comparable for minority and low-income students. Gains from participation in READ 180 were especially noticeable for African American students in the 1st year only, who

had gains on vocabulary, comprehension, spelling, and on the total reading assessment, with large effect sizes—.43 standard deviations for vocabulary, .30 for comprehension, and .63 for total reading. The 9.2 scale score gain on the spelling test reflected an effect size of better than one-quarter standard deviation improvement (.26) compared to the control group. The low-income students who participated in READ 180 had statistically significant gains on the vocabulary, comprehension, and total reading test in the 1st year only, with effect sizes of .3, .3, and .5 standard deviations respectively. There were no statistically significant impacts of READ 180 on the SAT10 for gender or the other race/ethnicity subgroups. There were no significant impact effects found in the other two studies for the subgroup analyses.

For summary purposes, findings are presented for the main student outcome measures only, which include DIBELS ORF, WJ-III, and SAT 10. Table 2 displays the impact findings by measure for each study.

Discussion

The main goal of the RCT studies was to examine whether structured reading curricula, adapted to afterschool settings, increased students' reading outcomes compared to the "business as usual" programs operating in the comparison afterschool classrooms. As part of a larger effort to examine the contribution of afterschool programs to improved student outcomes, information emerging from these trials was meant to add to the limited body of evidence on the impact of academically infused afterschool programs on student outcomes. Taken together, the overall implementation and impact findings documented in this set of studies indicate that well-implemented academic afterschool programs may have some impact on reading outcomes, and while some findings hold for low-income and minority student subgroups, more research is needed to replicate and extend these limited findings and examine their sustainability over time.

Table 2. Findings by Student-level Outcomes

	Adventure Island	Voyager Passport	READ 180 Year 1	READ 180 Year 2
DIBELS Oral Reading Fluency				
WJ-III - Letter-Word Identification				
WJ-III - Word Attack				
WJ-III - Passage Comprehension				
SAT 10 - Word Study Skills				
SAT 10 - Reading Vocabulary			+	
SAT 10 - Reading Comprehension			+	
SAT 10 - Spelling Assessment				
SAT 10 - Total Reading			+	

NOTE: Abbreviations of the findings are:
+: Finding of a positive impact
Blank cell: Finding of no impact

At the end of the 1st year, study results revealed improved reading outcomes for students participating in the READ 180 program. Although the impact of participation in READ 180 indicated a significant "bump" in reading outcomes, the changes made to the intervention and lack of findings at the end of the 2nd year rendered the combined impact across the 2 years difficult to interpret. The Adventure Island and Voyager Passport studies revealed that neither program had any impact on student reading outcomes after 1 year of implementation. Both studies encountered significant study and program implementation challenges that led to the reduction of their 2-year funding periods to 1-year time frames.

As in any rigorous research design applied to school settings, a number of issues arose across the studies that required program changes that may have limited the detection of program impacts. The challenges experienced during the implementation phases reinforce that rigorous efficacy research in applied settings requires an investment of significant monetary and human resources to ensure that programs are implemented fully and as designed, to fund the length of time (i.e., more than 1 year) instructors need to become proficient with program materials, and to collect information on the critical causal components operating in the contrast conditions. The differences between the causal components in the treatment condition and the control condition represent the relative strength of potential treatment impacts (Cordray & Pion, 2006), a critical feature of impact analysis but one that requires increased time and resources applied to the treatment contrast between conditions.

Consideration of Implementation and Impact Findings in RCTs

The research projects aimed to answer, using experimental methods, the extent to which structured reading interventions, implemented in afterschool settings, would positively affect reading skills in elementary grades, and whether those findings held for student subgroups. A secondary goal of the studies was to examine whether the quality of program implementation in the applied settings was related to student outcomes, using correlational methods. Although implementation findings typically are not treated as critical sources of information in RCT analyses, they are considered here in the discussion of the impact findings. More recently, the state of the field in RCT design in applied settings is calling for model specifications of and linkages between implementation fidelity and program impacts in the assessment of causality to improve post-experimental specification of interventions. The following discussion of implementation is referring generally to the installation of the structured reading programs in applied settings (Lipsey & Cordray, 2000).

Implementation fidelity may inform null impacts. The promising reading curricula evaluated in these studies were modified in response to various conditions that were occurring in the applied settings, a common response to on-the-ground problems that occur during efficacy trials. These studies were conducted with significant developer support to provide the best opportunity for on model program implementation. In two of the three studies, the modifications or "tolerable adaptations" (Cordray, 2008) to the programs were arrived at through consensus between afterschool program staff, evaluation researchers, and developers.

When programs undergo efficacy trials to verify their impact, assessment of implementation fidelity can be complicated by program changes, but the theoretical baselines against which fidelity is measured are potentially lost through the modifications. For example, at the beginning of the Adventure Island study modifications included a reduction in the number of days per week that the program was delivered without reducing the overall time per week that the students were exposed to the materials. The adaptations increased instructor buy-in because their concerns about weekly scheduling were accommodated, however, the modifications make it difficult to interpret precisely whether the implementation fidelity under the modified conditions ties back to originally proposed implementation targets when core components, such as number of days per week for delivery, may have impacted the overall delivery of the program. This example illustrates the usefulness of careful specification of program implementation impacts in efficacy study designs. According to some experts, implementation fidelity should be based on a priori expectations about the core components of the program, which are then measured for the faithfulness with which they are put in place during on-the-ground implementation; a well-designed analysis model should uncover the impact of the quality of

implementation fidelity, or faithfulness to the pre-stated intervention model, on the significant levels of program impact (Cordray, 2008; Dane & Schneider, 1998). With this level of clarity, the modifications to core program components and information gathered on implementation fidelity may provide meaning for the interpretation of null or limited impact findings and post-experimental intervention adjustments.

Instructor skill with program implementation may inform impacts. In general, instructors' skill with the use of program materials improved over the school year in these studies, as reflected in improvements on program implementation ratings assessed during afterschool classroom observations. As is common for professional support offered to teachers using structured programs, the 1st year sessions taught and supported the use of basic program routines and materials, troubleshot technical problems, and assisted instructors' understanding of student progression through levels of instruction. By the end of the school year, instructors mostly reached a mechanical level of expertise with the programs, but often had difficulty reaching this level of proficiency. It is reasonable to expect that in order to gain higher-than-basic levels of expertise with structured program materials instructors require more than 1 year of exposure to and practice with the program, especially in applied settings such as afterschool programs where there are fewer program days available in the school year compared to regular school classes. While no specific instructor threshold (i.e., amount of time needed to surpass basic expertise with the materials) was projected to benchmark proficiency, the fact that Adventure Island and Voyager Passport were implemented for only one year may have dampened program impacts due to the limited time instructors had with the materials. Additionally, greater attention to variation in instructor effects (e.g., experience, education, incentive pay) on implementation quality may have lent interpretive power to null or limited impact findings in these efficacy studies.

Exposure to the intervention and dosage thresholds may inform impacts. Implementation quality and "dosage," as measured by student attendance rates in the afterschool programs, were considered critical factors in whether programs led to improvement of student outcomes. Each program's potential positive impact on achievement assumes a full-year

of "on-model" implementation. Given the reduced amount of program time available in the afterschool programs, with typical late start and early end dates, the number of days available for instruction is far less than the optimal "dosage" days specified by the treatment curriculum developers. The reduction of time available to implement the program, particularly in the Adventure Island and Voyager Passport studies, along with the voluntary nature of afterschool attendance, limited the amount of time instructors had to gain confidence with the programs and the exposure that students had to the materials as instructors improved over time. In combination, these factors made it difficult to assess the impact of the reading programs on afterschool participants. While on the whole, instructors' ability to implement the programs improved over time, instructors may not have had enough experience with the materials to achieve the levels needed to fully activate the program components that would have contributed most to improved reading. Well-specified thresholds for dosage levels that allow the core components of the reading programs to "take effect" would increase the measurement precision of implementation elements and would assist a more meaningful interpretation of findings.

Discussion of overall limited program impacts. The afterschool RCT staff exerted an admirable amount of diligence toward study and program implementation, monitoring, and implementation and student assessment data collection. These studies were funded modestly, and as mentioned above, rigorous efficacy studies require significant resources to be well implemented. The projects recruited and maintained adequate sample sizes for the most part, but adequate sample sizes are especially important to attain when planning for the statistical power necessary to meet minimally detectable effect sizes to test student subgroups of interest (e.g., minority and low-income student subgroups), especially given the high attrition rates in afterschool programs. The limited program impacts found in this set of studies may be attributable to the 1-year time frames for two of the studies, and that all of the studies were conducted in challenging instructional conditions that required program modifications.

Overall, participation in READ 180 led to improved performance on reading assessments after 1 year

of implementation, but the null findings for the program effects in the 2nd year was surprising and left unanswered questions about overall impact. According to READ 180 literature, the program is designed for struggling readers, so the inclusion of students who were proficient in reading in the 2nd year sample may have interfered with the ability to find a difference between the treatment and the regular afterschool program. The importance of recruiting samples matched to the program's identified target population, while not always feasible, is highlighted by findings that are difficult to interpret in light of the significant impacts and large effect sizes after installation of only 1 year of the program.

Implications for rigorous research in applied settings. The difficulties associated with implementing the rigorous designs of these studies in applied settings may have contributed to the lack of findings across the studies. Many of the afterschool sites were inexperienced with providing structured, full-year afterschool programming, not to mention assimilating the demanding procedures associated with randomizing students, training instructors assigned specifically to treatment classrooms, and the related challenges of RCT designs (e.g., tracking and maintaining students in separate conditions, accommodating evaluation researchers needs for monitoring attendance and student assessments). Program requirements were difficult for many of the afterschool programs to accommodate given budget limitations and staffing problems. The research demands imposed on applied settings are often difficult for administrators and instructors to meet given the already compressed time and resources available in most afterschool settings. The added rigor of RCT designs requires that stakeholders are made fully aware of the research requirements and conditions and that their willingness to participate reflects not only that they understand these points but that their conditions (e.g., sample characteristics, staffing resources) are vetted for appropriateness given the specifics of the study.

Another unique contribution of this work to the field is the use of a consortium model to develop, guide, refine, and complete the RCT studies. This effort allowed for researchers and research organizations that are typically competitors to collaborate to improve individual work and to address common concerns that

emerged over the course of the trials collectively, bringing resources and expertise to bear on significant or anticipated issues. Within the context of the ARC, this project demonstrated the value of collaboration in maximizing the quality and quantity of the work as well as establishing a potentially valuable model for conducting rigorous trials within a research and development environment.

There are several limitations to the studies that should be mentioned. The short time span between the release of funds for the projects and project start up (approximately 3 months), while common, limited the amount of critical planning time to recruit and/or prepare sites, roll out professional development to sites, and ensure the fit between sites and program specifications were sound before the trials began. A related limitation was the length of the study implementation (1 year for two projects and 2 years for one project) and the potential restraint this short amount of time had on the program's ability to reach full implementation and detect changes in academic achievement outcomes among participants. All of the interventions were in an incipient phase of adaptation to afterschool settings, which implied that ongoing program adaptations were necessary. The success with which those adaptations could be made without negatively impacting the overall study design may have impacted program effects by the end of the studies. Given the challenges encountered with implementation, an additional limitation was that each study used a different implementation measure that varied in terms of observation units and fidelity measurement standards across studies. Finally, a caveat to consider with these studies is that because of the challenges encountered in measuring and reaching full implementation, it is difficult to conclude that the studies clearly articulate the degree to which highly structured reading curriculum can be beneficial in afterschool settings—because of this, there is still a critical need for efficacy trials to address this goal in the field of afterschool.

Conclusions

One of the key challenges to the success of these studies, and RCT studies in general, is the requirement of conducting large-scale research projects effectively with limited resources. The RCT studies experienced a number of challenges related to limited financial

and personnel resources and competing goals for afterschool services. The limited findings reflect that more research, funded at levels that allow for multiple years of implementation and the collection of large enough samples of students to ensure adequate statistical power, is needed to reach thresholds that meet significance levels for impact and subgroup analyses. A number of challenges encountered in this set of studies highlighted the need for improved measurement of implementation fidelity and the collection of information in contrast conditions to address treatment strength and improve the chances of detecting and interpreting program impacts.

Academic programming in afterschool settings remains a potentially rewarding and important endeavor. Gaining a better understanding of the relative effectiveness of the types of academic assistance offered in afterschool settings (i.e., unstructured and structured academic enrichment, tutoring, and homework assistance) and the target groups, organizations, and conditions that have the most impact remains an important focus for the field in terms of policy, research and practice. In addition, the studies in this report highlight the increasing importance of recognizing and becoming more sophisticated about the threshold of programming needed to attain impacts (dose-response) which in afterschool settings are inherently confounded by program fidelity, dose and duration challenges. Namely, many structured academically focused afterschool programming or curricula, like the ones in these trials, are adapted from curricula modeled in day-school settings and therefore have inherent challenges to fidelity when placed in an afterschool setting. Many programs operate only 4 days a week for 2–3 hours per day with about 45–60 minutes per day focused on academic enrichment, limiting the program treatment or dosage. In addition, most programs begin later in the school year and end before it does, limiting the duration of students' exposure to the program. Many important questions remain in this area. One important implication of this work is the continued need to support demonstration research and development efforts of instructional resources for core academic subjects that could be used in afterschool settings as has been attempted with these trials as well as other ongoing trials funded by IES (see Black, et. al., 2008).

The program specifications of most structured reading programs, common to interventions targeting low-achievement students, are resource intensive for the typical afterschool program, the program developer, and the research teams. While schools and districts often choose off-the-shelf programs with limited information about what would be appropriate for their contexts, the mismatch between what interventions usually require and the resources available in applied settings can contribute to a gap in understanding between all stakeholders regarding the potential costs and benefits of particular programs. Whether implementing reading curricula in afterschool programs stands to increase reading achievement as part of a research study or in practical application, the improvement of students' reading outcomes is a high-impact imperative that will benefit from additional research evidence gathered in applied settings.

References

August, D. & Shanahan, T. (2006). *Developing literacy in second language learners*. Report of the National Literacy Panel on Minority-Language Children and Youth. Wahwah, NJ: Lawrence Erlbaum.

Black, A. R., Doolittle, F., Zhu, P., Unterman, R., Grossman, J. B., (2008). *The evaluation of enhanced academic instruction in after-school programs: Findings after the first year of implementation* (NCEE 2008-4021). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Borman, G., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N. A., & Chambers, B. (2005). The national randomized field trial of Success for All: Second-year outcomes. *American Educational Research Journal 42*(4), 673–696.

Borman, G., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal, 44*(3), 701–731.

Calderón, M. & Slavin, R. (Eds.) (2001). *Effective programs for Latino students*. Mahwah, NJ: Erlbaum.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Lively, T. J., & White, C. E. (2004). Closing the Gap: Addressing the Vocabulary Needs of English-Language Learners in Bilingual and Mainstream Classrooms. *Reading Research Quarterly, 39*(2), 188–215.

Cordray, D. S. (2008, July). *Assessing interventions and control conditions in RCTs: Concepts and methods*. Presentation for the IES/NCER Summer Research Training Institute: Cluster-Randomized Trials. Evanston, IL.

Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin and P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103–124). Washington, D.C.: American Psychological Association.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 23–45.

Frechtling, J., Silverstein, G., & Zhang, X. (2003). *Evaluation of the Voyager Universal Literacy System.* Retrieved March 30, 2006, from www.voyagerlearning. com/ ResearchStudyDocuments/WashDC_CleOH_Westat_Yr1.pdf

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.

Hecht, S. A. (2003). *A study between Voyager and control schools in Orange County, Florida.* Retrieved March 30, 2006, from www.voyagerlearning.com/ResearchStudyDocuments/ OrangeCounty_FL_Treatment_Control_Study_2002-2003.pdf

Hecht, S. A. & Torgesen, J. K. (2002). *Within school treatment and control study.* Retrieved March 30, 2006, from www.voyagerlearning.com/Research StudyDocuments/ OrangeCounty_FL_Treatment_Control_Study_2001-2002.pdf

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

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

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: Reports of the subgroups.* Bethesda, MD: National Institute of Child Health and Human Development.

National Reading Panel. (2001). *Put reading first: The research building blocks for teaching children to read. Kindergarten through grade 3.* Bethesda, MD: National Institute of Child Health and Human Development.

National Research Council. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Roberts, G. & Allen, A. S. (2003). *Impact of the Voyager Universal Literacy System as measured by PALS in Virginia.* Retrieved April 3, 2006 from www.voyagerlearning.com/ ResearchStudyDocuments/ULS_measuredby_PALS_ Richmond_ VA.pdf

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd Ed.). Boston: Allyn & Bacon.

Slavin, R. E., & Madden, N. A. (Eds.) (2001). *One million children: Success for All.* Thousand Oaks, CA: Corwin.

Slavin, R. E., Groff, C. & Lake, C. (2008). Effective reading programs for middle and high schools: A best evidence synthesis. *Reading Research Quarterly, 43*(3), 290–322.

Vandell, D., Reisner, E., & Pierce, K. (2007). *Outcomes linked to high-quality afterschool programs: Longitudinal findings from the study of promising afterschool programs.* Irvine, CA: University of California and Washington, DC: Policy Studies Associates, Inc.

AFTERSCHOOL RESEARCH CONSORTIUM

Advancing Research, Improving Education

SEDL

Available online at www.sedl.org/afterschool/resources/