
Results of a Randomized Experiment of the Mondo Publishing's *Bookshop* Literacy Program

Henry May

Jonathan Supovitz

University of Pennsylvania¹

June 21, 2011

¹ This report contains data collected for an independent study conducted by Jonathan Supovitz and Henry May.
The opinions expressed are those of the author and do not reflect the positions of the University of Pennsylvania in any way.

Executive Summary

Mondo Publishing's *Bookshop* Reading Program is a comprehensive literacy program for Grades K-5. This report presents the results of a randomized experiment of the impact of the *Bookshop* Program that was implemented in the New Haven, CT school district in Grades K-3 only. For this experiment, 16 schools were matched based upon prior achievement and student demographics to produce 8 pairs. One school from each pair was then randomly assigned to implement the Mondo *Bookshop* program. Experimental designs like this are considered by researchers to be the strongest approach to determine program effectiveness because they eliminate threats to study validity, and produce causal estimates of program effects. The experimental research design used in this study meets the standards of evidence established by the What Works Clearinghouse (<http://www.whatworks.ed.gov/>) and the criteria for scientifically-based research as defined by No Child Left Behind.

In this study, program impacts on 3rd grade state reading test scores were examined for three cohorts of students. Impacts were estimated for students at the end of 3rd grade, and persistence of the impacts was examined through the fifth grade, two years after students stopped participating in the *Bookshop* program. The results indicate that in all three cohorts, students who participated in *Bookshop* had better reading performance than students in the control group, with impact estimates for two of the three cohorts surpassing the criterion for statistical significance. The effect sizes associated with these differences were substantial; students who participated in the *Bookshop* Program made gains in 3rd grade that were 21 to 27 percent larger than those of students in the comparison group. After students stopped participating in the *Bookshop* program, they continued to make substantial annual gains; however, the students who received only one year of the treatment (i.e., Cohort 1) lost their advantage in reading performance relative to students in the control group after they left the program. This suggests that the *Bookshop* program improves students reading achievement considerably in even just one year of treatment, but unless students participate in *Bookshop* for more than one year, the additional gains achieved with the *Bookshop* program may be lost when students return to traditional reading instruction.

Overview

Mondo Publishing's *Bookshop* Reading Program is a comprehensive literacy program for the elementary grades. This report presents the results of a randomized experiment of the impact of the *Bookshop* Program that was implemented in the New Haven, CT school district in the early elementary grades only (i.e., K-3). In this study, eligible schools were randomly assigned to implement the *Bookshop* program or to serve as a control school. Randomized experimental designs like this are considered by researchers to be the strongest approach to determine program effectiveness because they eliminate threats to study validity and produce causal estimates of program effects.

In this study, program impacts on students' state reading test scores are examined for three cohorts. Impacts were estimated for students in 3rd grade, and persistence of the impacts was examined through the fifth grade, two years after students stopped participating in the *Bookshop* program. The results indicate that in all three cohorts, students who participated in *Bookshop* had better 3rd grade reading performance than students in the control group, with impact estimates for two of the three cohorts exceeding the criterion for statistical significance. Results also suggest that multiple years of participation in *Bookshop* may be necessary to ensure sustained effects of the program.

Program

Bookshop Reading, a core reading program for Grades K–5, is taught during a daily literacy block, which may include a 30-minute Skills Block, followed by a 60-minute Reading Block. Whole group instructional strategies include shared reading and read aloud. Small group strategies may include oral language-reading, phonics, shared reading, guided reading, and reciprocal reading.

Differentiated instruction is a critical component of *Bookshop*. Teachers are expected to guide individual students along comprehensive instructional pathways based on assessed readiness needs, as opposed to having all students reading the same core selection (which may have been adapted to accommodate both greater and lesser needs). Multiple copies of precisely leveled texts across a broad range of levels are available to support instruction at a student's exact point of need. The core objective is for students to apply and practice critical comprehension strategies.

Each *Bookshop* grade-level program is built on a scope and sequence that spirals instruction through a continual loop of new and reviewed phonics skills and comprehension strategies. The program is designed to support teachers in their use of assessment data to form small groups and identify each group's starting point on the scope and sequence. Lesson plans model focused teaching, which incorporates clear comprehension focuses and explicit teacher language.

Specific Program Aspects

Assessment: *Bookshop* promotes ongoing formative assessment. The data generated are intended to provide a measure of student progress and information to guide modification of instruction yet to come. This allows teachers to both monitor individual progress and to differentiate instruction. Mondo's multiple formative assessment measures provide a comprehensive diagnostic profile of each student. The resulting data are designed to improve targeted instruction, which begins precisely at point of need and proceeds at an appropriate pace. To do so, Mondo assessments establish starting points for instruction, measure individual performance and growth over time, and support teachers as they implement effective classroom practices.

Oral Language: Research has repeatedly documented the strong link between a student's limited oral language proficiency and a subsequent struggle with reading comprehension (Burns & Helman, 2008; Resnick & Snow, 2009; Wasik, 2010). Many students arrive at school with limited proficiency in—or understanding of—instructional language. They have difficulty reading grade-level texts that include language structures beyond their capacity. The Mondo Oral Language Assessment identifies oral language proficiency and possible risk of reading failure. *Bookshop's* oral language-reading lessons are designed to explicitly model, build, and scaffold the language structures students need for reading comprehension—the critical “missing link” between social and instructional language.

Phonics: *Bookshop Phonics* is a research-based program in which instruction is intended to be systematic and explicit. A field-tested assessment determines three levels of intensity for daily differentiated instruction, resulting in pre-teaching for students needing extra skills support, and appropriate small-group placement based on what students already know. The research-based scope and sequence moves through a continuum of new and reviewed phonemic awareness, phonics, and word recognition skills. Built-in progress monitoring enables teachers to measure response and fine-tune instruction.

Professional Development

Mondo Professional Development is designed to support and sustain the improvement of a school district's literacy instruction. A primary goal is to help leadership (district and school) and teachers learn how to use data to drive teaching and learning. This is Mondo's cornerstone for school-based, professional learning driven toward improving literacy outcomes. The expectation is that eventually, the use of data becomes embedded in everyday practice as principals, coaches, and teachers engage in discussions about data and their implications for teaching and learning. Typical professional development partnerships last approximately three years.

Another goal is a focus on instructional practices linked to students' specific literacy data. Mondo believes that in order to make an impact at the classroom level, educational leaders must establish conditions that provide teachers with adequate ongoing support to enable this shift in practice (e.g., weekly team meetings, literacy coach support, defined literacy block timeframes). Consequently, Mondo Professional Development emphasizes the use and understanding of formative assessment data to reveal the changes needed in teaching practices and materials.

Data

The study was conducted using data from 16 elementary schools in New Haven, Connecticut. The New Haven Public School district has a total of 50 schools and serves about 21,000 students. Just over half the students in the district are African American, approximately one-third are Hispanic, and just over 10 percent are white. From the district's 30 elementary schools, 16 were chosen to participate in the study. These 16 schools were then grouped into matched pairs of schools with similar poverty rates (expressed as the percent of students receiving free/reduced lunch assistance) and prior (2005-06) performance. From each pair of matched schools, one was randomly assigned to receive the *Bookshop* Program, and the other served as a control school. This type of randomized experimental design using matched pairs improves pre-intervention equivalence of treatment and control groups, and enhances statistical power to detect program effects.

Although schools were matched on demographics, there may still be differences in the demographics of individual students in the *Bookshop* and comparison schools. For this reason, we also present the performance and demographic information for individual students. Table 1 shows pre-intervention comparisons of student performance and demographics prior to participation in the *Bookshop* program. The data indicate very similar baseline performance at the 3rd, 4th, and 5th grade levels. The demographics between the two groups are also very similar, with the only slight difference being a slightly higher percentage of white students in the *Bookshop* schools (9 percent compared to 6 percent).

Although *Bookshop* is a K-5 program, the roll-out of the program in the study district was staged over time, beginning with grades K-3 in the 2006-07 school year, then expanding into the upper elementary grades three years later, in 2009-10. This means that the three cohorts of students in this study each received different amounts of exposure to the *Bookshop* treatment. The first cohort participated in *Bookshop* for only one year, while they were in 3rd grade during the 2006-07 school year. The second cohort participated in *Bookshop* for only two years: first when they were in 2nd grade in 2006-07 and then in 3rd grade in 2007-08. The third cohort participated in *Bookshop* for three years: first when they were in 1st grade in 2006-07, then in 2nd grade in 2007-08, and lastly in 3rd grade in 2008-09.

The *Bookshop* program in New Haven elementary schools was not rolled out to 4th and 5th grades until after the conclusion of this study. Therefore, the analysis of data beyond 3rd grade in this study does not test the impact of the program during its implementation. Instead, the analysis of 4th and 5th grade data in this study allows us to study whether the program effects achieved during the implementation of the program (i.e., through 3rd grade) were maintained after students left the program and returned to regular reading instruction.

Table 1. Descriptive Statistics Comparing Baseline Performance and Demographics of Students in Treatment and Control groups

	Pre-Experimental Cohort	
	Control	Treatment
Number of Students	538	564
3rd Grade Reading Score		
Mean	376	375
Standard Deviation	46	45
4th Grade Reading Score		
Mean	403	401
Standard Deviation	44	43
5th Grade Reading Score		
Mean	434	430
Standard Deviation	49	43
Percent Male	50.1%	49.6%
Percent African American	84.6%	87.4%
Percent Asian	11.9%	10.1%
Percent Hispanic	10.6%	10.8%
Percent Native American	<0.1%	<0.1%
Percent White	6.0%	9.0%
Percent Free/Reduced Lunch	83.0%	83.0%
Percent ELL	12.0%	11.0%
Percent Special Education	10.0%	11.0%

Methods

Impact of the *Bookshop* program was estimated by comparing the 3rd through 5th grade reading achievement trajectories of students in the treatment and control groups. By comparing students' achievement trajectories, we can estimate the impact of the program at the end of 3rd grade (the final year of the program), and we can estimate the sustained impact of the program on the reading performance of students in the first two cohorts after they leave the program and move into traditional reading instruction in 4th and then 5th grade.

Reading achievement scores from grades 3, 4, and 5 on the Connecticut Mastery Test (CMT) were obtained for all students enrolled in 3rd grade in treatment and control schools during

the 2005-06 through 2008-09 school years. Therefore, we have 3rd, 4th, and 5th grade reading scores for students in the pre-experiment cohort (in 3rd grade during 2005-06) and the first treatment/control cohort (in 3rd grade during 2006-07); we have 3rd and 4th grade reading scores for students in the second cohort; and we have just 3rd grade scores for students in the third experimental cohort.

The statistical model used to test the significance of program effects was a Hierarchical Linear Growth Model (HLGM). This model allows us to (a) evaluate pre-treatment differences between treatment and control schools, (b) estimate impacts of *Bookshop* on students' performance in 3rd grade, (c) estimate impacts of *Bookshop* on students' annual gains after 3rd grade, and (d) control for student background characteristics including gender, race, free/reduced lunch eligibility, English language learner status, and special education status. The HLGM included three levels: test scores at level 1, nested within students at level 2, who were nested within schools at Level 3. Fixed effects were included for cohorts, and additional random effects for baseline achievement and growth were included at Level 3 for the experimental blocks (i.e., the matched pairs of treatment and control schools).

Because only three years of data were available, the growth curve trajectories were modeled as linear, with an intercept parameter representing performance in 3rd grade and a slope parameter representing annual gains after 3rd grade. These two parameters were then modeled at Level 2 as a function of student characteristics, and then at Level 3 as a function of cohort and treatment assignment status. Treatment by cohort interaction parameters were included at Level 3 in order to produce separate impact estimates for each experimental cohort. Using the notation of Raudenbush & Bryk (2002), the mathematical form of this three-level hierarchical linear growth curve model is:

$$\text{Level 1 (test scores): } Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{Grade}) + \varepsilon_{tij}$$

$$\text{Level 2 (students): } \pi_{0ij} = \beta_{00j} + \sum \beta_{10k}(\text{Student Characteristic}_k) + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + \sum \beta_{11k}(\text{Student Characteristic}_k) + r_{1ij}$$

$$\text{Level 3 (schools): } \beta_{00j} = \gamma_{000} + \gamma_{001}(\text{Treatment}) + \sum_{c=1}^a \varphi_{00c}(\text{Cohort}_c)$$

$$+ \sum_{c=1}^a \theta_{00c}(\text{Cohort}_c \times \text{Treatment})$$

$$+ \text{Block}_{00j} + \omega_{00j}$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}(\text{Treatment}) + \sum_{c=1}^2 \varphi_{10c}(\text{Cohort}_c)$$

$$+ \sum_{c=1}^2 \theta_{10c}(\text{Cohort}_c \times \text{Treatment})$$

$$+ \text{Block}_{10j} + \omega_{10j}$$

The Level-1 equation specifies the trajectory of each student’s growth curve, with π_0 representing 3rd grade achievement and π_1 representing growth after 3rd grade, along with the variation (represented by ϵ) of the test scores around the latent growth trajectory. The Level-2 equations show how various student characteristics are related to students’ 3rd grade achievement (i.e., π_0) and growth after 3rd grade (i.e., π_1). The Level-3 equations show treatment effects on 3rd grade achievement (i.e., γ_{001}) and growth after 3rd grade (i.e., γ_{101}) and how these effects differ for each cohort (i.e., θ), along with additional parameters to test for baseline differences between cohorts (i.e., ϕ).

Results

Before presenting the results of our statistical models, we provide the unadjusted descriptive statistics for students in the treatment and control groups as they matriculate from 3rd to 5th grades. The cohort columns in Table 2 show the performance of the three cohorts of students as they move through the grades. These represent the raw performance of students in the three cohorts as they proceed from 3rd to 4th to 5th grades. These unadjusted data provide a picture of comparative student performance as they move from 3rd to 4th to 5th grades and presage differences in performance in subsequent models which control for prior achievement and demographic differences between treatment and control students.

Table 2. Descriptive Statistics Comparing Unadjusted Performance of Cohorts of Students in Treatment and Control Groups.

	Cohort 1		Cohort 2		Cohort 3	
	Control	Treatment	Control	Treatment	Control	Treatment
Number of Students	473	558	404	514	384	440
3rd Grade Reading Score						
Mean	377	381	387	384	383	389
Standard Deviation	47	43	46	44	43	44
4th Grade Reading Score						
Mean	412	409	425	419	n/a	n/a
Standard Deviation	49	48	39	43	n/a	n/a
5th Grade Reading Score						
Mean	449	438	n/a	n/a	n/a	n/a
Standard Deviation	45	41	n/a	n/a	n/a	n/a

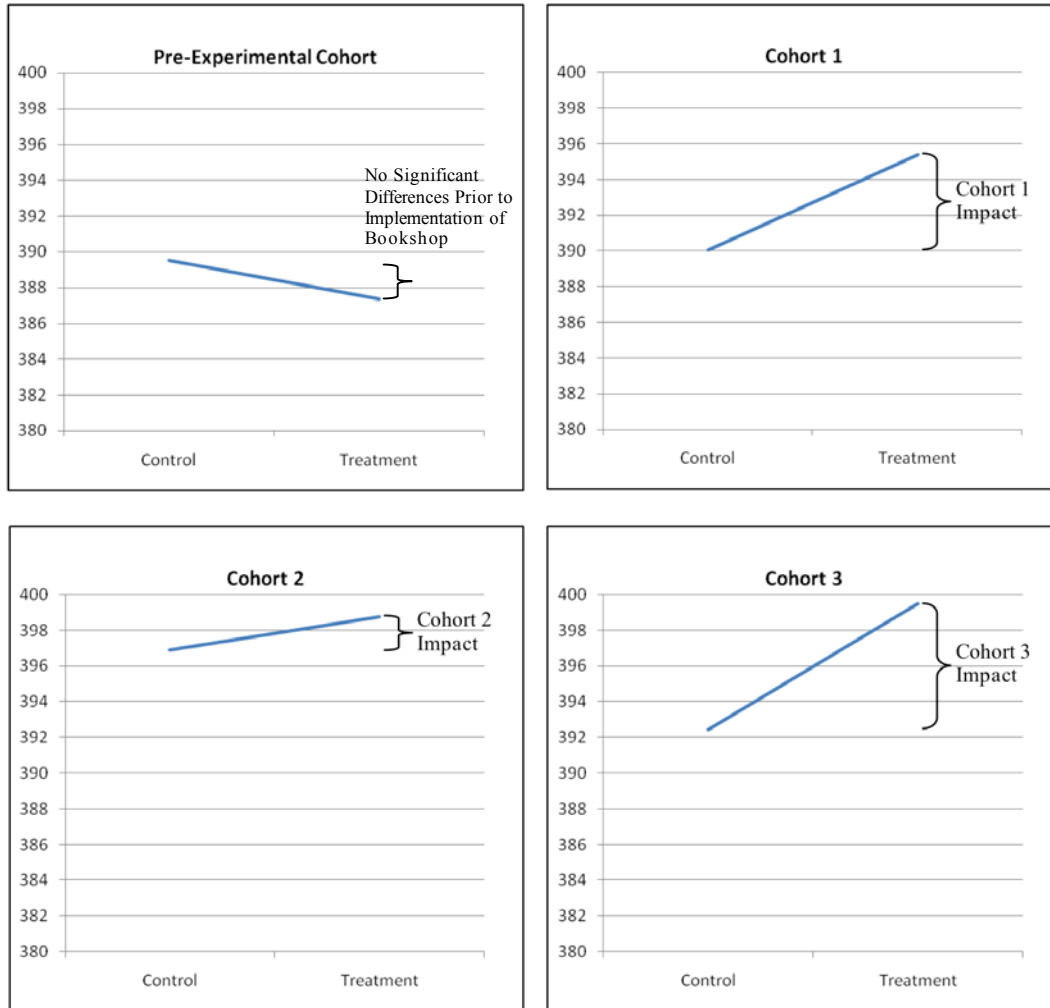
Note: n/a = not administered (e.g., Cohort 2 entered 5th grade after the conclusion of this study)

Results from the HLGGM impact model are shown in Table 3 (see page 8). First, there were no detectable pre-experimental differences between the treatment and control groups in terms of 3rd grade achievement ($\gamma_{001} = -2.14$, $p = .64$) or growth after 3rd grade ($\gamma_{101} = 1.31$, $p = .41$), suggesting that the random assignment produced comparable treatment and control groups prior to the intervention. Significant positive effects on 3rd grade reading scores are evident for two of the three cohorts. In Cohort 1, the *Bookshop* treatment group scored 5.36 points (.12 standard deviations) above the control group ($\theta_{001} = 5.36$, $p = .08$); and in cohort 3, the *Bookshop* treatment group scored 7.06 points (.16 standard deviations) above the control group ($\theta_{001} = 7.06$, $p = .04$). In Cohort 2, both the treatment and control groups scored significantly higher than the baseline performance ($\phi_{002} = 7.41$, $p < .01$), and the *Bookshop* treatment group scored an additional 1.86 points (.04 standard deviations) above the control group, but this difference is not statistically significant ($\theta_{001} = 1.86$, $p = .56$). In Cohorts 2 and 3 there were no differences in performance after the students left the *Bookshop* program and continued on in grades 4 and 5. For Cohort 1, which received only 1 year of the program while in 3rd grade, there was a reduction in growth after termination of the *Bookshop* treatment ($\theta_{101} = -5.14$, $p < .01$), which was almost exactly opposite of the initial positive impact, suggesting that gains made by cohort 1 as a result of the treatment were lost after the treatment was discontinued.

Interpretation of the 3rd grade impact estimates presented above is enhanced when compared against the expected annual growth rate for control group students. As shown in Table 3 (see page 11), the baseline growth rate after 3rd grade is 25.71 points per year ($\gamma_{100} = 25.71$, $p < .001$). By dividing each raw effect estimate by 25.71 and multiplying by 100%, the impact estimate is expressed as the percentage gained above and beyond the average annual gain for control group students. For Cohort 1, the treatment effect of 5.36 points translates to a gain for treatment students in the 3rd grade that was 21% larger than the gains for control students. For Cohort 3, the treatment effect of 7.06 points translates to a gain for treatment students in the 3rd grade that was 27% larger than the gains for control students.

Figure 1 (see page 10) presents a visual depiction of the adjusted 3rd grade treatment/control differences for the non-experimental cohort and the three treatment cohorts. The estimates from this graph are calculated based on the results of the HLGGM. For the pre-experimental cohort, performance is slightly higher in the control group, but this difference is not statistically significant. In Cohorts 1 and 3, there is a clear advantage for the treatment group, and the differences are statistically significant. In Cohort 2, both the treatment and control groups have high average scores, and there is a slight advantage for the treatment group, but this difference is not statistically significant.

Figure 1. Mondo Bookshop Program Impacts Across Three Experimental Cohorts



It is also interesting to note that the baseline annual growth rate for students in this study is 26 points per year ($\gamma_{100} = 25.71$, $p < .001$), which is not statistically significantly different from the statewide average growth from 3rd to 4th grade of 28 points. Thus, the annual growth for the study district was similar to the statewide average. Given that students in this district experience greater disadvantage (e.g., higher percentage of free lunch eligibility, ELL status, and special education status) than the rest of the state, these annual gains are quite remarkable.

Table 3. Parameter Estimates of Program Impact from Hierarchical Linear Growth Model

	Third Grade Reading	Growth After Third Grade
Intercept (Pre-Experimental Baseline) (γ_{000} , γ_{100})	389.50**	25.71***
Pre-Experimental Differences (γ_{001} , γ_{101})	-2.14	1.31
Treatment Effect for Cohort 1 (θ_{001} , θ_{101})	5.36~	-5.14***
Treatment Effect for Cohort 2 (θ_{002} , θ_{102})	1.86	-3.16
Treatment Effect for Cohort 3 (θ_{003} , θ_{103})	7.06*	n/a
Cohort 1 vs. Baseline (φ_{001} , φ_{101})	0.54	1.09
Cohort 2 vs. Baseline (φ_{002} , φ_{102})	7.41**	1.97
Cohort 3 vs. Baseline (φ_{003} , φ_{103})	2.94	n/a
Male (β_{101} , β_{111})	-10.54***	2.97***
Race/Ethnicity (African American Reference)		
Asian (β_{102} , β_{112})	29.73***	9.21**
Hispanic (β_{103} , β_{113})	6.08***	0.39
Native American (β_{104} , β_{114})	32.17	-9.74
White (β_{105} , β_{115})	21.70***	-1.35
Free/Reduced Lunch Status (β_{106} , β_{116})	-4.68***	0.84
English Language Learner Status (β_{106} , β_{116})	-18.86***	0.85
Special Education Status (β_{107} , β_{117})	-43.35***	-0.18
Variance of Random Effects		
Block Effect ($Block_{00}$, $Block_{10}$)	42.46	3.16
School Level Variance (ω_{00} , ω_{10})	71.71	6.18
Student Level Variance (r_0 , r_1)	1131.22	8.78
Residual Variance (ϵ)	336.71	n/a

Note: ~p<.10, *p<.05, **p<.01, ***p<.001

Growth after 3rd grade for Cohort 3 is “n/a” given only one year of available post-treatment data.

Summary and Discussion

The results from this randomized experiment show significant positive impacts of the impact of the *Bookshop* Reading Program on New Haven students' third-grade reading scores on the state assessment. The experimental design used in this study yielded treatment and control groups that were not significantly different immediately prior to the implementation of the *Bookshop* program. Therefore, subsequent differences can be attributed to the causal effects of the program. In other words, student performance would not have been as high if the *Bookshop* program had not been implemented.

Program impacts indicate that in all three cohorts, students who participated in *Bookshop* had better 3rd grade reading performance than students in the control group, with impact estimates for two of the three cohorts surpassing the criterion for statistical significance. The effect sizes associated with these differences were substantial; students who participated in the *Bookshop* Program made gains in 3rd grade that were 21 to 27 percent larger than those of students in the comparison group. After students stopped participating in the *Bookshop* program, they continued to make substantial annual gains; however, the students who received only one year of the treatment (i.e., Cohort 1) lost their advantage in reading performance relative to students in the control group after they left the program. This suggests that the *Bookshop* program improves students reading achievement considerably in even just one year of treatment, but unless students participate in *Bookshop* for more than one year, the additional gains achieved with the *Bookshop* program may be lost when students return to traditional reading instruction.

Furthermore, given that students in the study district were already making substantial annual gains without the program, the sizable impact of *Bookshop* is especially promising. For the students in this study, the average annual gain exceeded the baseline annual gain by as much as an additional 27%. This suggests that prior to the implementation of *Bookshop*, students in these New Haven elementary schools the study district were progressing at a rate that was similar to the statewide average. After the implementation of *Bookshop*, students in the treatment schools experienced gains that exceeded the statewide average by about 20%. As the program is expanded to upper elementary grades, persistence of these effects has the potential to dramatically reduce performance gaps in reading achievement for students in this high-needs district.

References

Burns, M. K. & Helman, L. A. (2008). What does oral language have to do with it? Helping young English-language learners acquire a sight word vocabulary. *The Reading Teacher*, 62(1), 14-19.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Resnick, L. B. & Snow, C. E. (2009). *Speaking and Listening for Preschool through Third Grade*. Pittsburgh, PA: University of Pittsburgh; Washington, DC: The National Center on Education and the Economy.

Wasik, B. A. (2010). What Teachers Can Do to Promote Preschoolers' Vocabulary Development: Strategies From an Effective Language and Literacy Professional Development Coaching Model. *The Reading Teacher*, 63(8), 621-633.