Evaluation of Literacy Program: Project Transformation

Evaluation of Literacy Program: Project Transformation

Pooja Shivraj, M.S., Ph.D.

Founder and Executive Director of Research, Evaluation, and Analysis
Shivraj Consulting, LLC

September 2016

Contact Information

Pooja Shivraj, M.S., Ph.D.
Founder and Executive Director of Research, Evaluation, and Analysis
Shivraj Consulting, LLC
Address: PO Box 1441, Coppell TX 75019
Email: pooja@shivrajconsulting.com
Web: http://www.shivrajconsulting.com

Evaluation of Literacy Program: Project Transformation

Introduction

The purpose of this report is to describe the evaluation of a literacy program developed by Project Transformation. Within the report, we present a brief introduction of Project Transformation and their mission, the rationale for the design of the program, the evaluation design, the methodology and analyses used for evaluation, and the highlights of the conclusions obtained from the results. Limitations from the study are provided along with implications for the program and recommendations for the future.

Project Transformation: Mission and Background

Project Transformation is a Christian-based nonprofit founded in 1998 under the auspices of the North Texas Conference of the United Methodist Church. In October of 1997, Sarah Wilke and Dr. Leighton Farrell came together to imagine a solution to address three main needs: (a) socially-conscious leadership training and vocational exploration for college-aged young adults interested in pursuing ministry, (b) academic, health, spiritual, and recreational programming for children and youth in the North Texas area, and (c) revitalization of inner-city United Methodist churches. In essence, the mission of Project Transformation is to engage young adults in purposeful leadership and ministry, support underserved children and families, and connect churches to communities in need.

One of the primary needs and challenges which Project Transformation's participant constituency faces are summer reading loss. Research shows that meeting this key need as early as possible has positive effects on children and youth's cognitive and non-cognitive growth and development.

Project Transformation prioritizes marginalized children who qualify for government assistance as participants in its programs. At Project Transformation North Texas, upwards of 85 percent of children self-define as a minority race or ethnic group. Unfortunately, statistics show that poverty and race frequently go hand-in-hand. This marginalized population of children traditionally experiences tremendous literacy challenges and significant summer reading loss. While a host of interrelated factors contribute to the racial and socioeconomic disparities in student achievement, it is well documented that children and youth living in poverty and low-income communities perform far below their counterparts in middle- and upper-class communities.

Next, we provide a brief literature review on the rationale and importance of the literacy program developed by Project Transformation.

Literature Review

A 2011 research report by the RAND Corporation (McCombs et al., 2011) found that summer learning loss, particularly in reading, disproportionately affects economically disadvantaged students. Most disturbing is the fact that the summer learning loss is cumulative, so that over time these different learning rates between lower- and upper-income students contribute substantially to the achievement gap, a gap that cannot be made up during the school year. A 2007 National Center for Summer Learning (NCSL) research report found that two-thirds of the ninth-grade reading achievement gap between low- and higher-income students can be attributed to unequal access to summer learning opportunities (Alexander, Entwisle, & Olson, 2007).

As the income gap has widened, so has the achievement gap. According to a 2011 study, the achievement gap, which has grown significantly for at least three decades, is 30 to 40 percent larger among children born in 2001 than among those born 25 years earlier (Reardon, 2011). According to a 2009 report by McKinsey & Company, impoverished students (defined as those receiving federally subsidized lunches) lag academically behind their more affluent counterparts by roughly two years. The poverty gap appears early and persists over the student's lifetime. Schools composed mostly of low-income students perform much worse than schools with more affluent enrollments (McKinsey & Company, 2009).

It is also well documented that poverty and proficiency in reading by the end of third grade is critical to a child's educational success and likelihood of graduating high school (Hernandez, 2011). Based on the 2011 National Assessment of Educational Progress report, for low-income children, four out of every five children score less than "proficient" in reading assessments. A report by the Annie E. Casey Foundation (2010) found that students who are not reading proficiently by the end of 3rd grade are four times more likely to not graduate high school on time. Yet, America's literacy rates, especially among low-income children, remain a national challenge. Based on the 2011 National Assessment of Educational Progress, more than two-thirds of the nation's fourth graders are not "proficient" readers, including 34% who scored below "basic." In 2012, the National Center for Education reported that for low-income children, more than four out of every five scored less than "proficient," including 54% who scored below "basic." Seventy-five percent of students with literacy problems in the third grade will still experience literacy difficulties in the ninth grade (Alliance for Excellent Education, 2003).

A significant achievement gap also exists between English language learners (ELL) and non-ELLs at all socioeconomic levels, but this gap is most severe for students at the lowest socioeconomic levels (Garcia & Frede, 2010). The number of students entering U.S. schools speaking little or no English has grown exponentially over the last decade. In a 2011 report from National Clearinghouse for English Language Acquisition, from 1998 to 2009, the number of ELLs enrolled in public schools increased from 3.5 million to 5.3 million. The vast majority in (60%) of the low-income students served by Project Transformation North Texas fall into the ELL category.

Access to high-quality affordable summer and after-school learning opportunities is among the most crucial issues that face all low-income families, and no less so in the communities Project Transformation serves. Numerous national studies consistently show that children who are at home without adult supervision for some part of the day, especially after school until a parent

3

returns from work, also known as "latchkey kids, " are more likely to be involved in violent crime, substance abuse, tobacco use, and other high-risk behaviors. Despite a growing awareness that summer learning loss is a major contributor to the achievement gap, the number of children participating in summer enrichment programs is startlingly low. According to the National Conference of State Legislatures (NCSL), only 25 percent of children (approximately 14.3 million) participate in summer learning programs. While there are many high-quality summer learning programs across the country, there are simply not enough to meet the growing demand. Based on NCSL parent intern surveys, 56 percent of non-participating children would likely participate in a summer learning program, and nearly half of those children who are likely to participate are low-income students.

Addressing the academic "summer slide" for low-income students remains a significant problem. Summer learning loss, particularly in reading, is cumulative. Over time these different learning rates contribute substantially to the achievement gap, which cannot be made up during the school year. A 2007 NCSL research report found that two-thirds of the ninth-grade reading achievement gap between low- and higher-income students can be attributed to unequal access to summer learning opportunities. By the end of 5th grade, low-income students are up to 2.5 years behind in reading compared with their higher-income peers, largely because low-income students' reading skills stagnate or decline over the summer break. However, recent research from the RAND Corporation has demonstrated that quality summer learning programs can prevent summer learning loss and even boost student achievement (McCombs et al., 2011).

Literacy Program Design

Project Transformation delivers quality reading intervention over the 8 weeks of summer based on 17 years of proven best practices and aligned with national standards. The literacy program achieves its outcomes through daily one-to-one reading with community volunteers. Reading progression at Project Transformation is truly a community effort. The program comprises an elementary program for students in grades 1-5, and a youth program for students in grades 6-9. Both programs have a reading coordinator who trains volunteers, coordinates assessments, pairs children and volunteers, and tracks outcomes. Children are pre- and post-assessed during the first and last weeks of the program using Guided Reading Levels determined using the running records of the reading coordinator and trained volunteers.

Elementary children read one-to-one each morning with a community volunteer from local community organizations, churches, and high school students. The morning elementary literacy component focuses on building fluency, comprehension, and self-efficacy as children read aloud then reflect on the story with a volunteer. This component lasts for 30 minutes and is for 4 times a week during the 8 weeks of summer. In the afternoon, children interact with young adults in small groups according to their skill level. This also occurs for 30 minutes 4 times a week during the 8 weeks of summer.

In the youth literacy program (Grades 6-9), young adults work with youth to host read alouds, monitor youth self-guided reading, lead daily journal entries, and oversee youth comprehension projects. It is also divided into a morning and afternoon component. The individual, self-guided reading occurs in the morning for 30 minutes, three times a week over the 8 weeks of summer.

Group reading and projects also occur in the morning for 30 minutes once a week during the 8 weeks of summer.

Logic Model

The logic model for the program is shown in Table 1.

Table 1. *Logic model for the reading intervention*

Population Needs	Inputs	Outputs	Outcomes
Lack of out-of-school summer supervision Summer learning loss Lack of reading skill	Volunteers Administrative and program support staff Host church sites AmeriCorps members Intervention (Reading program)	Trained program support staff for administration of program Consistent implementation of program across sites	A statistically significant increase in reading level from pre- to post-assessment at each grade

Purpose of the Study

The purpose of this study was to evaluate the program developed by Program Transformation, by determining the statistical and practical significance of the growth of students from pre- to post-summer after receiving the program, if any.

Methods

Proposed Evaluation Designs

The first proposed evaluation design involved a quasi-experimental pre-test post-test design with a control and treatment group. Given that recruitment of estimated sample sizes were not possible, a nonequivalent dependent variables (NEDV) design (Shadish, Cook, & Campbell, 2002), where a single group of participants were delivered *two* pretests and *two* posttests, was proposed. One set of pre- and post-assessments were designed to be the Reading Assessments, and the other set of pre- and post-assessments were designed to be math assessments, used as covariates for reading. A depiction of this design in shown below.

Pretest (O_1) ---------------Treatment (X) ------------Posttest (O_1)

Pretest (O_2) ---Posttest (O_2)

Here, O_1 depict the Reading Assessment, while O_2 depicts the math assessment used. The hypothesis was that the treatment (literacy program) will only impact O_1 and not O_2, since the program was not designed to impact O_2. Here, O_2 acts similar to a control group and models what would have likely happened if the literacy program had not been delivered to the sample. Hypothetically, if the literacy program had worked, the slope of O_1 would be greater than the slope of O_2, with maturation over time. Math was chosen as the covariate given the high correlation shown in the literature between math and reading across grade levels (e.g., Crawford, Tindal, & Stieber, 2007; Jiban & Deno, 2007; Hecht, Torgesen, Wagner, & Rashotte, 2001; Lerkkanen, Rasku-Puttonen, Aunola, & Nurmi, 2005). For example, analyses of student performance on the Programme for International Student Assessment (PISA; Kelly et al., 2013), known to have a particularly high reading demand in its mathematics items, shows a correlation of 0.95 between PISA mathematics country mean scores and PISA reading country mean scores (Wu, 2010).

However, for the implementation of this design, a sufficient sample that took all four assessments (pre- and post-reading, and pre- and post-math) was not obtained. Table 2 shows the sample that took the reading assessment, and the sub-sample that took the math assessments.

Table 2. *Sample sizes obtained*

Grade	Only pre-reading assessment	Both pre-and post-reading assessment	All: pre- and post-reading *and* pre- and post- math assessment
Grade 1	97	65	7
Grade 2	117	79	18
Grade 3	126	87	29
Grade 4	102	64	13
Grade 5	102	64	9
Grade 6	83	58	NA
Grade 7	54	40	NA
Grade 8	59	47	NA
Grade 9	31	22	NA

Implemented Evaluation Design

Due to the lack of attainment of sample, the one-group pretest-posttest design was implemented, where all participants in the group were tested prior to the program and then tested after the program. A depiction of this design is shown below.

Pretest (O_1) ---------------Treatment (X) ------------Posttest (O_1)

Guided Reading Levels from O_1, the Reading Assessment, are used as the dependent variable, and compared to the Posttest for statistical significance.

Sample

The final sample that was used for this design is shown in the second column of Table 1 (Sample that took both pre- and post-reading). This sample comprised of 50.3% Female and 45.8% Male (with 3.9% unreported). The sample was also broken down into 60.1% Hispanic or Latino, 14.7% Black or African American, 9.9% White, 1.5% Asian, .4% Two or More Races, and .2% American Indian or Alaskan Native (with 13.2% unreported.

Instrument: Reading Assessments

Reading Assessments has three main goals: (a) to understand whether a child is reading above, at, or below their current grade level; (b) to identify children who are not reading yet; and (c) to determine the appropriate grade level each child is capable of reading at.

Reading Assessments is intended to be delivered in the first and last weeks of the summer for Grades 1 through 9, before and after the literacy program is implemented. Passages for this program differ from the beginning to the end of the summer program such that students will never read the same passage twice. Passages are leveled from reading level A through Z, with A being the easiest and Z being the hardest. An example of a beginner reading level student passage is shown in Figure 1.

There are certain basics for the delivery of Reading Assessments. Once a quiet location is found for delivering these assessments, the grade level correlation chart is used to determine which passage is to be used to start with for the child. For example, a grade 1 child would start at

reading level A, compared to a grade 2 child who would start at a reading level F. Testers have two binders in possession, one comprising student passages, and the other with tester's own running records. As children read the passage, testers follow along on their running record, crossing out words that children miss and counting the number of errors. Testers also note when children self-correct so that it does not count as an error. Errors include things such as omitting a word, substituting another word in place of the word in text, inserting an extra word, or pausing for longer than 5 seconds to the told the word in text. Mispronouncing a name or saying a word wrong prior to self-correcting do not count as errors. Testers pay close attention as children read so they are able to ask questions about the passage to check for comprehension. An example of a tester's running record is provided in Figure 2.

Students read passages up until their reading level is determined. Testers keep a Reading Assessment Log to keep track of the assessments they have done. Once a child's reading level is determined, their reading letter (or level) is circled and the number of errors is written in the row next to the child's name. Passages do not have to be read in one sitting and are not timed.

Instrument: Math Assessment

Scores from a math pre- and post- assessment was attempted to be used as a covariate. However, given the lack of participation in these assessments, the use of the covariate was dropped from the research design. The instrument and procedures for delivery will hence not be described for purposes of this paper.

Procedures for Delivery and Implementation

The program was delivered with fidelity between the beginning of June and the end of July 2016 at nine site locations and by a total of twenty testers.

Analyses

Given the ordinal nature of the outcome variable, the reading assessment scores, a Wilcoxon Signed Rank Test (by grade) was used to determine differences in pre- and post-test scores. Reading levels were converted to numeric for purposes of analyses and significance testing (i.e., A=1, B=2, etc.). The effect size for the test was calculated by dividing the z-estimate by the square root of the number of observations over the two time points. Cohen's (1988) criteria of 0.1, 0.3, and 0.5 for small, medium, and large, respectively, were used for interpretation of effect sizes. All analyses were conducted in R (R Core Team, 2016).

Results

Table 3 shows the sample sizes, the mean pre- and post-reading levels, and the effect sizes by grade level. A mean post-reading level of 5.01 for Grade 1, for example, corresponds to a reading level of about E averaged across 65 participants. The % of students who increased by at least one reading level is recorded, along with the % of students who decreased by one or more reading levels. The % of students who stayed at the same level can be determined by subtracting the percentages from 100. Figures 3 shows the growth using boxplots for each of these grades.

Table 3. *Reading Assessment evaluation*

Grade	Sample size	Mean pre-reading level (SD)	Mean post-reading level (SD)	% of students increased	% of students decreased	Effect size
Grade 1	65	3.75 (3.72)	5.01 (4.48)	63.1%	4.6%	.18
Grade 2	79	7.92 (4.83)	10.04 (4.87)	79.7%	6.3%	.20
Grade 3	87	10.34 (5.57)	12.07 (5.57)	72.4%	4.6%	.15
Grade 4	64	15.80 (4.40)	17.41 (4.40)	76.6%	4.7%	.18
Grade 5	64	16.64 (4.55)	18.67 (4.55)	76.6%	0.0%	.18
Grade 6	58	18.02 (3.04)	20.00 (3.04)	86.2%	3.4%	.14
Grade 7	40	18.95 (4.08)	20.27 (4.08)	80.0%	12.5%	.17
Grade 8	47	21.30 (3.31)	22.53 (3.31)	72.3%	6.4%	.15
Grade 9	22	22.95 (2.62)	23.86 (2.62)	68.2%	13.6%	.23 (NSS)

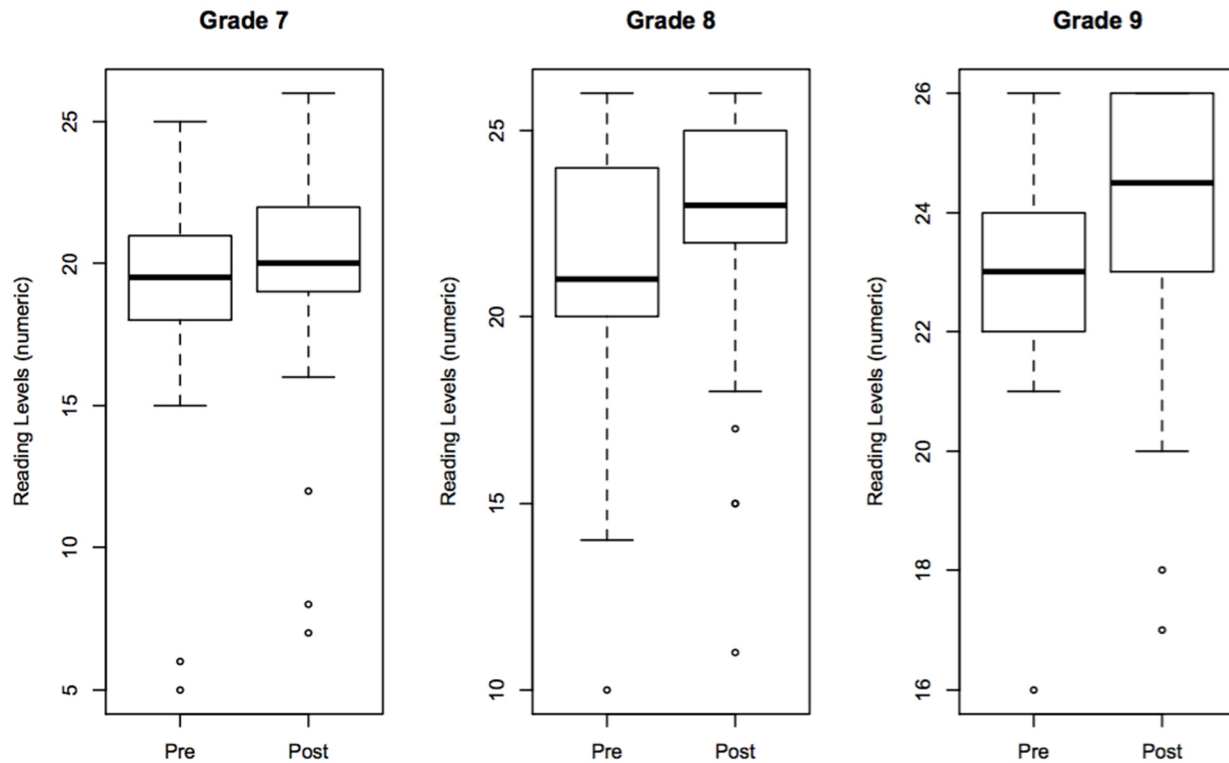*Note: NSS = not statistically significant*
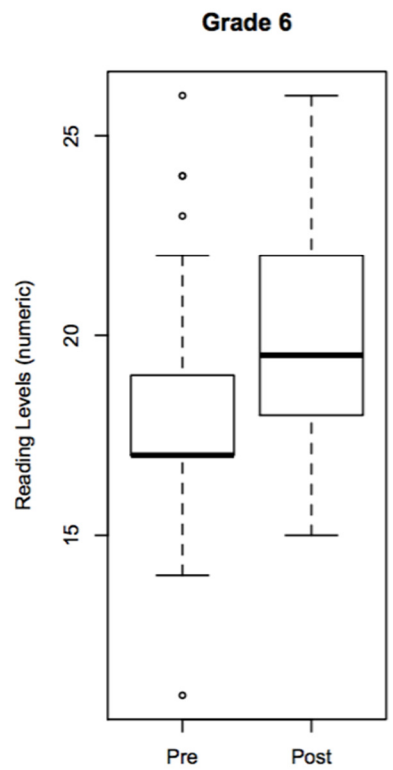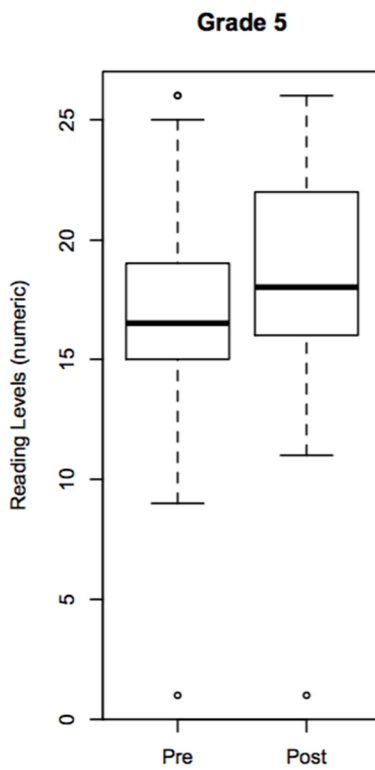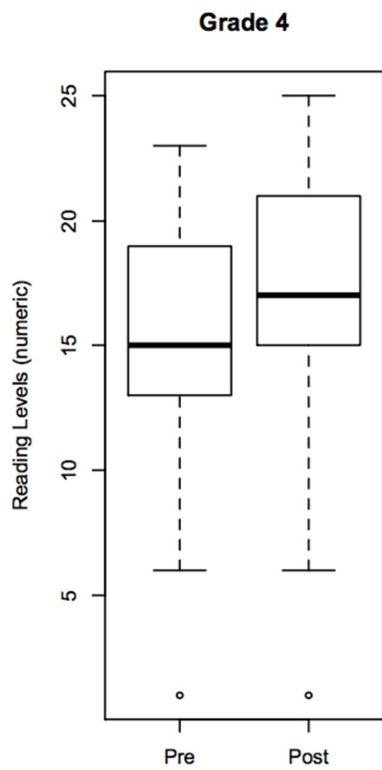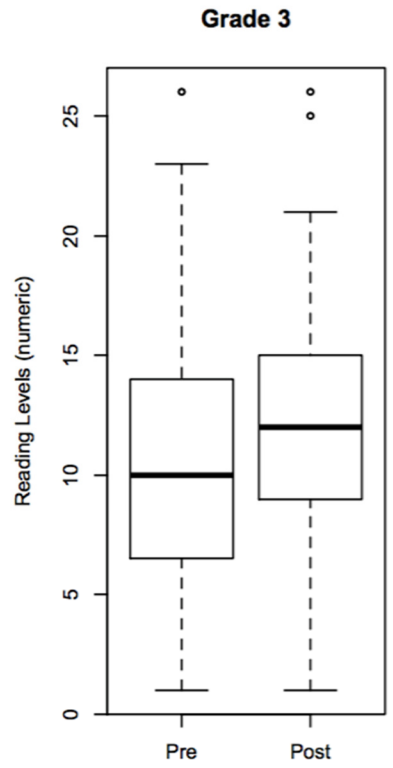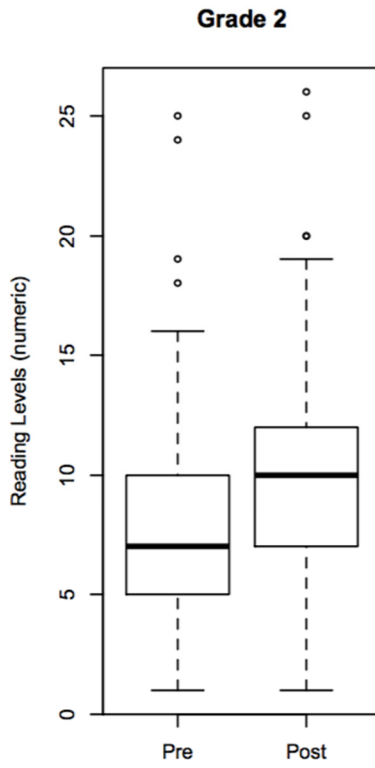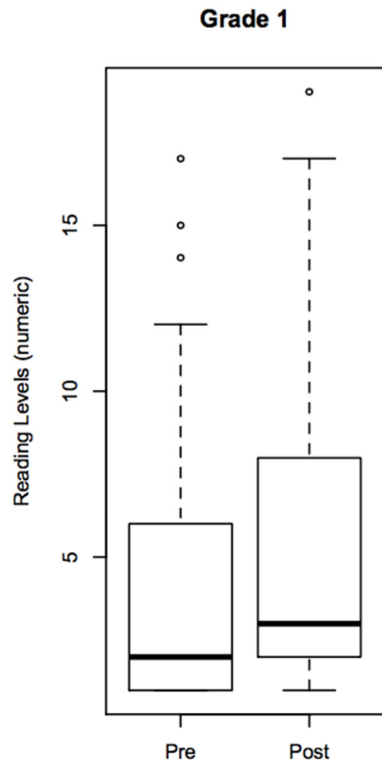
Figure 3. *Visual representation of growth measured using Reading Levels from pre- to post-intervention.*

Conclusions

Summary of Overall Findings

At all grade levels from Grade 1 through Grade 8, there was statistically significant growth from pre- to post-test. The effect size calculated at every grade level was small (Cohen, 1988). The growth in Grade 9 was not statistically significant; however, the effect size was small. This may be attributable to the small sample size.

Limitations to interpretations

While the design was intended to incorporate a covariate to strengthen the evaluation, the sample size deemed necessary for the evaluation was not obtained. The results obtained from this study are nevertheless promising; however, there are multiple factors to consider while interpreting these results, as with any study with a similar research design structure (Shadish et al., 2002).

- Threat of history or maturation: It is impossible to say if this same sample of children would have had the same growth outcome regardless of the intervention. It is possible that children in the community watch a TV show or read books at home that enable this type of growth over time. However, given the research on summer slump (Alexander et

al., 2007; McCombs et al., 2011), this seems unlikely. Also, given the short time span between the pre- and post- assessment, threat to history seems unlikely.

- Mortality threat: The attrition of sample seems pretty significant from Table 2. However, a mortality threat is not common with a one-group pre-posttest design since there is only one group and the comparison is between their pre- and post-scores. Table 4, however, shows some significant differences by grade in the original sample versus the final sample, which could potentially bias the results and interpretation if regression of the kids with lower scores who remained in the sample (and hypothetically, regression of kids with higher scores who dropped out of the sample as shown in Table 4) occurred.

Table 4. *Attrition from original sample to final sample – differences in mean pre-reading scores*

Grade	Original sample size	Mean pre-reading level (SD)	Final sample size	Mean pre-reading level (SD)
Grade 1	97	6.97 (4.07)	65	3.75 (3.72)
Grade 2	117	11.36 (5.50)	79	7.92 (4.83)
Grade 3	126	14.58 (6.31)	87	10.34 (5.57)
Grade 4	102	19.84 (5.51)	64	15.80 (4.40)
Grade 5	102	22.30 (5.81)	64	16.64 (4.55)
Grade 6	83	23.43 (4.00)	58	18.02 (3.04)
Grade 7	54	25.37 (4.88)	40	18.95 (4.08)
Grade 8	59	28.32 (4.38)	47	21.30 (3.31)
Grade 9	31	30.58 (2.91)	22	22.95 (2.62)

- Generalizability: Generalizability of the sample is always a question for any study and needs to be considered when looking in to extending the results of programs based on smaller samples recruited based on convenience sampling.

Future Directions and Recommendations for Future Research

The results are extremely promising. However, multiple recommendations for future program evaluation are recommended before modifications to the program are made or the program is scaled.

- **Testing different components of the program:** The program was evaluated purely on a global level of growth from pre- to post- intervention on the Guided Reading Level. It would be optimal to evaluate each component of the program – phonemic awareness, fluency, etc. – to determine areas of strengths and weaknesses.

- **Comparison with a control group for effects:** It would be optimal to use a control group (or a covariate, less ideally) to evaluate the program for the impact of the program.

- **Disaggregating results by demographics:** With a larger sample, it would be optimal to determine stability by subgroup – i.e., differences in growth by gender, race/ethnicity, ELL, etc. While this information was available for this study, sample sizes rendered the

assumptions for the analyses challenging to meet (i.e., assumptions of normality within groups and/or assumptions of near-equal samples within subgroups).

References

Alexander, K., Entwisle, D., & Olson, L. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72,* 167-180.

Alliance for Excellent Education. (2003). *Adolescents in crisis.* Paper presented at the first annual American High School Policy Conference. Washington, DC.

Annie E. Casey Foundation (2010). *Early Warning! Why Reading by the End of Third Grade Matters.* KIDS COUNT Special Report: Author.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Routledge.

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*, *7*(4), 303-323.

Garcia, E. E., & Frede, E. C. (2010). Young English Language Learners: Current research and emerging directions for practice and policy. New York, NY: Teachers College Press.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*, 192-227.

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* The Annie E. Casey Foundation.

Jiban, C. L., & Deno, S. L. (2007). Using math and reading curriculum-based measurements to predict state mathematics test performance are simple one-minute measures technically adequate? *Assessment for Effective Intervention*, *32*(2), 78-89.

Lerkkanen, M. K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J. E. (2005). Mathematical performance predicts progress in reading comprehension among 7-year olds. *European journal of psychology of education*, *20*(2), 121-137.

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

McKinsey & Company. (2009). The economic impact of the achievement gap in American schools: Summary of Findings. Author.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

14

Reardon, S. F. (2011). The widening academic achievement gap between rich and poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–115). New York: Russell Sage Foundation.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference.* Boston: Houghton Mifflin.

Wu, M. (2010). *Comparing the similarities and differences of PISA 2003 and TIMSS* (OECD Working Papers No. 32). Paris: OECD Publishing.