America Reads – Mississippi:

An impact evaluation of the program's 2015-2016 school year

AMERICA
READS
MISSISSIPPI

Tutors With A Mission

Prepared by:

terraluna
COLLABORATIVE

Impact Evaluation of the America Reads - Mississippi Program

November 2016

Authors

This study was conducted by evaluators from the TerraLuna Collaborative:

Jason R. Altman, M.A.

Juan M. D'Brot, Ph.D.

Nadia M. Akhtar, M.A.

This report represents the work and perspectives of the authors and is the product of a contracted external impact evaluation. It does not represent the position or opinions of CNCS, the federal government, or the programmers.

Acknowledgements

Citation

Altman, J. R., D'Brot, J. & Akhtar, N. (2016). America Reads – Mississippi: An impact evaluation of the program's 2015-2016 school year. TerraLuna Collaborative: Minneapolis, MN

EXECUTIVE SUMMARY

America Reads – Mississippi (ARM) is a national service AmeriCorps program funded through a grant from the Corporation for National and Community Service (CNCS). **ARM is dedicated to improving the reading skills of students, encouraging public awareness and support of literacy, and helping to increase the number of certified teachers in Mississippi.** ARM AmeriCorps members strive daily to achieve this mission, primarily through tutoring K - 3rd grade students in reading both one-on-one and in small-group settings during school, and in extended-day/-year programs.

SPECIFICALLY, THIS STUDY INVESTIGATED TWO RESEARCH QUESTIONS:

First, what was the America Reads - Mississippi program's impact on improving student academic gains in reading during the 2015-2016 school year? *Second, what was the impact of the same program on student attitudes towards academic and recreational reading?* The findings related to this second question are of particular interest, as there has been little previous study of the effect of tutoring on student attitudes.

Due to limited resources, tutoring interventions using paid staff are typically not available for all struggling readers at all grade levels within all buildings in a state. However, the feasibility of administering a state-wide program utilizing AmeriCorps volunteer member tutors is presented in the full report. Policymakers may refer to the findings when deciding how to distribute resources to students struggling in reading.

The impact evaluation discussed in this report occurred during the second year of a three-year evaluation cycle that supported program function starting with the 2014-2015 school year. The story of this three-year journey is presented in the pages that follow with significant details regarding the 2015-2016 school year impact evaluation, and appendices providing even more extensive information to interested readers.

The evaluation design consisted of a randomized-controlled trial in some grades and a comparison to a matched sample in others. A grand total of 1,141 K-3 students attending 20 schools located across the state of Mississippi participated in the impact evaluation. Many of these schools had received a state grade of a D or F by state auditors related to their core reading instruction program. Even more, all involved students were considered at risk of falling behind grade level expectations during the 2015-2016 school year.

Selected students met with AmeriCorps members, who had received a small amount of program-sponsored training, at their school, outside of the traditional 90-minute reading block. For this reason, programming was meant to support and not supplant whole-class instruction. Program goals called for two hours (4 days of 30 minutes) of tutoring per week. Across all schools, the mean duration of intervention per student was nearly 23 weeks, and nearly 48 hours of tutoring.

Members were supported by a site supervisor at their location and two regional administrators. They communicated detailed information regarding tutoring sessions and student outcomes with these administrators as well as the America Reads - Mississippi state office located at the Institute for Higher Learning in Jackson. Program implementation fidelity varied across locations and across members. These factors were considered in secondary analyses.

Results of the evaluation were inconclusive at primary and secondary levels of analysis. Though analysts uncovered positive changes for some grades and some students related to either student attitudes or academic achievement, these changes failed to show statistical significance. Further, secondary analyses indicated that a high number of other independent variables influenced final outcomes, and that the unique impact of increased intervention magnitude was typically over-shadowed by the common impacts of groups of these independent variables. For example, student attitudes towards reading were found to be positively impacted by total intervention minutes, however the common or shared impact of intervention minutes and observed implementation adherence was negative.

In an effort to place these findings in the context of previous evaluations of similar programs, we returned to the literature and published findings in this topic area. Please note that readers should refer to the narrative present in the body of this report for full citations of those whose ideas are represented in this executive summary.

First, our failure to generate positive significant effects for student achievement was seen across a number of different studies, especially when instrumentation measured higher-order thinking such as comprehension rather than basic skills. In fact, authors inferred from previous findings that the failure of students to progress in a group tutoring setting, or even in a one-on-one setting, likely indicates a need for more individualized attention from a professional educator. At the early elementary grades assessed in this study, that could mean beginning conversations regarding an Individualized Education Plan (IEP).

We also measured variable implementation fidelity across locations, and even within locations among individual members. This variation often took the form of varied supportive infrastructure, such as having a unique classroom to conduct tutoring in or having limited opportunities to meet with students for planned sessions. A number of other authors also noted these limitations and added that, even in instances where tutors used a standardized curriculum with students, member capacity to impart prescribed interventions varied widely.

Third, previous authors provided commentary specific to the conversation surrounding the use of a standardized curriculum in tutoring sessions and the impact on implementation fidelity when a day-to-day alignment to the classroom is desired instead. The America Reads - Mississippi program relied on the close interaction between teachers, site supervisors, and tutors in cooperative curricular decisions. Our monitoring of implementation at the local level during the 2015-2016 school year informed us that these interactions were not always present in quantity or quality of interaction to fully support informed curricular alignment by tutors. In fact, other authors have argued that while such a collaborative curricular effort is desirable in theory, in practice, it is unlikely that other school personnel will have or make the time necessary to personalize lessons for each individual child.

Finally, our study's findings related to student attitudes should also be considered in the context of previous research. Though scant previous inquiry exists, one study produced similar results through the evaluation of a tutoring program in one large urban community. In fact, not only did the author find a significant positive impact for only one subgroup of students on just one of the attitude scales, but follow-up evaluation one year later showed no significant continued effect on student attitudes. The findings of that study and this evaluation also provide evidence of a lack in connection between student attitudes and student achievement. This signals the need to consider the full complexity of the relationship between an intervention like tutoring with individual student attitudes and student achievement, which echoes the work of literacy theorists.

It is important that the reader be prepared to reconsider the counterfactual claim that an intervention's merit or worth be wholly connected to it producing large effect sizes and statistically significant change as compared to a control. As authors are beginning to suggest, the successful implementation of research to practice efforts in previous decades has resulted in a strong and positive change in the rigor and effectiveness of the general education reading curriculum, and improvements in multi-tiered structures of support for students in need across schools. It is to be expected that students receiving regular school support without access to the individual intervention in question (a volunteer tutor) would still make sizable gains towards positive student outcomes across the span of a school year. Although the ideal scenario is to have a program with statistically significant improvement patterns when compared to a control, a program that can provide similar results to other interventions but that does so at the fraction of a cost while using volunteer interventionists with limited resources devoted to training could be considered to have proven its value as a sustainable alternative for schools with deficient resources.

As such, the **America Reads - Mississippi program** administrators continue to collaborate with institutions of higher education, local school districts, and local communities in light of these findings and the review of literature. Program refinement for the 2016-2017 school year included a strengthening of application materials and more hours and improved planning for pre-assignment training. At the same time, the program has used existing tools and has introduced new tools designed to support the mission of each of its members at the local level. These include a guide for members to use during their weekly meetings with school staff and a monthly electronic newsletter with links to resources for members. Finally, administrators have distributed some standardized curricular resources for use with students, including phonics flash cards, phonemic awareness sound box kids, alphabet kits for letter recognition, and alphabet letter tiles.

This study was limited by a number of factors and its findings should be considered in light of this information. To begin, secondary analyses indicated that implementation scores were a particularly strong indicator of positive change in student outcomes. However, as mentioned previously, implementation fidelity was varied across locations and even within locations among different AmeriCorps members. As a result, it is clear that readers should mark the important distinction between our investigation of the program "as implemented by local school personnel" and the program design itself.

Second, a number of the tools used to collect data for use in secondary analyses were developed during the course of the school year without the time needed for pilot, analysis, and modification. Even existing tools, when studied with a critical lens, did not

meet established research expectations for rigor, validity, and reliability. For this reason, it is important that the program continue to refine these tools in an effort to alleviate the challenge faced by program administrators in knowing and understanding adherence to program design at individual program sites across the state.

Finally, a number of methodological limitations presented themselves during the course of the study. Among them was the use of two different assessment tools for student achievement in 1st grade, which presented a challenge related to maintaining statistical power due to a lower sample size than expected for each of two groups. In addition, the partnership between buildings and program administrators did not provide for the opportunity for pre-intervention testing of student attitudes, and then later in the school year for only those students participating in the program. This made it impossible to measure student change for this outcome or the comparison of intervention student attitudes to those students not receiving tutoring. Instead, the evaluation team was required to investigate this outcome through non-experimental means without known baseline equivalence or a control group.

Perhaps most importantly, the theory behind the America Reads - Mississippi program includes outcomes at multiple levels. Though the data collected for this study focused on the student-level outcomes of achievement and attitude towards reading, the program also aspires to positively impact its members and its communities. For example, the program sees itself as a training ground with the capacity to entice potential educators to finish the required training in order to become licensed professionals. In addition, the program enlists its members in many community-based functions aimed at revitalizing awareness and engagement towards literacy-related issues. This study did not collect data to analyze impacts related to these other levels of outcomes.

In conclusion, readers should understand these key messages. First, the program findings, though not significant and positive, match inconclusive findings reported by others, especially in investigation of higher-order reading skills. Second, low-cost volunteer programming can serve as a valuable alternative to traditional intervention methods in certain contexts. And third, the evaluation process has already led to programming changes designed to maximize the impact of the America Reads – Mississippi program. Other programmers and stakeholders should consider these findings in the context of their own work, and potential evaluation efforts.

TABLE OF CONTENTS

INTRODUCTION

America Reads – Mississippi (ARM) is a national service AmeriCorps program dedicated to improving the reading skills of students, encouraging public awareness and support of literacy, and helping to increase the number of certified teachers in Mississippi. ARM AmeriCorps members strive daily to achieve this mission primarily through tutoring K - 3rd grade students in reading both one-on-one and in small-group settings during school and in extended-day/year programs.

ARM members provide tutoring to students with the greatest academic need in reading through implementing school-selected, research-based interventions and regularly assessing student gains and query for additional needs. ARM members also meet with the classroom teachers and school reading specialists to report on progress and inquire about promising future strategies on a student-by-student basis.

The program's design and implementation have been found to align with a sizable body of research related to the use of volunteers to provide a tutoring intervention to elementary school children. Though these studies sometimes differ on the use of curricula, tutor training techniques and magnitude, and specific settings, there are numerous individual studies as well as a collection of meta-analyses from which the program draws its vision.

Following a brief review of this literature, this report will outline the methods used to investigate impacts of the ARM program and will include a discussion of results and inferences related to findings. Specifically, this study investigated two research questions. First, what was the America Reads - Mississippi program's impact on improving student academic gains in reading during the 2015-2016 school year? Second, what was the impact of the same program on student attitudes towards academic and recreational reading?

This impact evaluation occurred during the second year of a three-year evaluation cycle that supported program function starting with the 2014-2015 school year. The details of this three-year journey are highlighted in Appendix A. However, on a broader note, the program spent the first year working closely with a consultant to prepare for an evaluation, including strengthening partnerships at the local level and initiating data use agreements. The second year consisted of the collection of student, member, and school data related to program implementation and student progress on the outcomes mentioned above. The third year consisted of the analysis of data and reporting of results in conjunction with an external evaluator hired to support in years two and three.

Review of Literature

Historically, studies of the reading performance of American children have found that a significant segment of youth have performed behind their peers in skills and ability. For example, at the turn of the 21st century, more than 2.8 million children were identified as requiring special services to remediate reading difficulties (U.S. Department of Education, 2000). Later, a study of the 2007 National Assessment of Educational Progress (NAEP), which samples students in various grade levels across all states in the country, found that 33% of fourth graders read below the basic level for their grade, and 67% of fourth graders read below the proficient level for their grade (J. Lee, Grigg, Donahue, & others, 2007). Later, in 2013, a report suggested that only 35% of American 4th-graders scored as proficient on the NAEP reading section (National Center for Education & Statistics, 2013), while the average scale score for the subject had improved less than 2% since 2002.

Many of these students who are performing behind their peers may also possess one or more risk factors that predict less positive future academic outcomes. For example, a study of disaggregated 2007 NAEP data showed that minority and low-income students have consistently underperformed compared to their White and middle- or high-income peers (J. Lee et al., 2007; Vanneman, Hamilton, Anderson, & Rahman, 2009). Studies that have used other instruments to measure student achievement concur with those findings seen on the NAEP. For example, Hernandez (2011) found that of children who have lived in poverty and were not proficient readers by third grade, 26% did not finish high school—six times the non-completion rate for proficient readers. Hernandez also found that this rate was highest for poor students who identify as African American (31%) and Hispanic (33%), lagging far behind their peers who identify as White (22%).

No matter the particular issue, affliction, or risk factor facing a particular child, the literature has shown that those who struggle with early reading skills will practice their reading skills significantly less than other students (Allington, 1994). The lack of practice for these students decreases their exposure to books and written information in the future (Stanovich, Cunningham, & West, 1998). Thus, a continuous cycle is initiated as the gap between good and poor readers widens over the elementary school years (Stanovich, 1986). This is a cycle that is very difficult to break, as evidenced by a recent study linking early childhood literacy skills and 8th-grade NAEP scores. Results of that study showed that not meeting a 1st grade, 3rd grade, or 5th grade reading benchmark resulted in a less than 50% chance of reaching proficiency in 8th-grade (Dogan, Ogut, & Kim, 2015).

This gap between advantaged students and minority and/or low-income students has been sustained, and even widened, in many areas, despite concentrated efforts towards helping minority children of poverty ("The Education Trust—West Response to the 2009 Nation's Report Card," 2010). At the same time, these students face challenges in more than the subject area of reading, as reading skills are necessary across many school subjects. According to Morris, Tyner, and Perney (2000), reading provides access to much of what is considered important in the school curriculum across a variety of subject areas; students who are below grade level early in elementary school will likely face an achievement gap in all subject areas in later grades.

In addition to poor academic outcomes, research has shown that failing to acquire literacy skills in the early elementary grades is associated with increased problem behaviors in school, higher probability of dropping out of school (as mentioned previously, Hernandez, 2011), limited employment opportunities, and a higher likelihood of living in poverty (Hitchcock, Prater, & Dowrick, 2004).

Research has shown that most reading difficulties are associated with core deficits in phonological processing affecting reading acquisition (i.e. phonological awareness), and the encoding of phonological information in memory (Adams & Gathercole, 2009; Snow, Burns, Griffin, & others, 1998). Hart & Risley (2003) found that for students from economically disadvantaged families, difficulties in reading are compounded by poor vocabulary and limited background knowledge. It is now understood that vocabulary is a direct contributor to reading comprehension and can operate as an additional risk factor above and beyond any issues with word recognition/decoding and language comprehension (Braze et al., 2016).

There is also research that has identified other reading skills that are indicators of future success. In the early grades, several meta-analyses have confirmed that foundational skills such as phonological awareness, phonics and word recognition, reading fluency, and attention to higher-order instruction in language and comprehension help students learn to read (Lonigan & Shanahan, 2009).

However, traditional efforts to target some of these skills with direct instruction or school-wide programming have often produced mixed results and no significant gains in targeted skills. For example, an evaluation of Reading First—once the nation's most comprehensive evidence-based reading curriculum and professional development for teachers—found that the intervention did not lead to statistically significant impacts on reading comprehension for students in first through third grades (Gamse, Jacob, Horst, Boulay, & Unlu, 2008). As a result, there is a documented need for additional interventions that focus on students with below-grade-level reading skills.
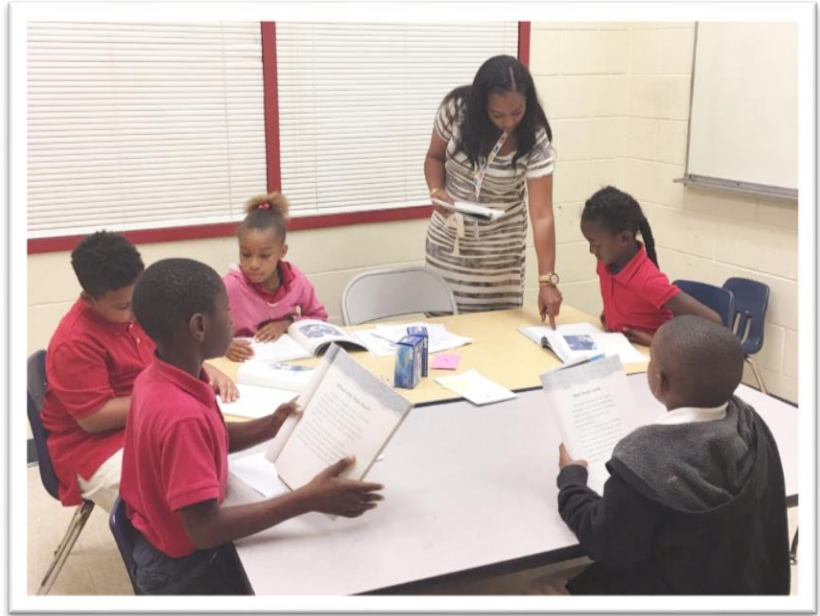
Tutoring

The positive effects of tutoring on reading skills have long been documented through literature reviews (Shanahan, 1998), program evaluations (Dorn & Allen, 1995), and meta-analyses (Cohen, Kulik, & Kulik, 1982; Elbaum, Vaughn, Tejero Hughes, & Watson Moody, 2000; Mathes & Fuchs, 1994; Wanzek et al., 2015; Wasik, 1998; Wasik & Slavin, 1993).

In a meta-analysis of 29 studies, Elbaum et al. (2000) found that tutoring interventions supporting students struggling in reading, when added to the general curriculum, had a positive impact on students. The author found that compared with control students, reading outcomes for tutored students had a mean weighted effect size of 0.41 across all studies. Later, Marulis and Neuman (2013) reviewed 51 studies with 138 effect sizes.

They found that word-learning interventions specifically contributed a mean effect size of nearly 1.0, indicating a strong positive impact. However, they also found that children with lower economic status gained significantly less from word-learning interventions.

It should be noted that in this review, we will often reference the effect size contributed by an intervention as reported by study authors. These results, as well as other key study details, are available in a table synthesizing the works most congruent with the ARM program's vision and policies in Appendix B. For those readers unfamiliar with the term "effect size," please know that the term references a standardized indication of impact so that study results can be compared to each other. In other words, the effect size allows us to compare the strength of the intervention across studies that use a myriad of instruments with different scales and reference points. Of note, an effect size above 0.4 is above average for educational research, and a value between 0.5 and 1.0 is roughly equivalent to about a school year's worth of educational growth for a student (Hattie, 2008).

Recently, Wanzek and co-authors (2016) completed a meta-analysis of studies of reading interventions for struggling readers from Kindergarten to third grade. They found a mix of studies that utilized tutoring interventions. The mean effect size reported for studies with tutoring interventions was about 0.5. The study did not find any moderating effects due to number of students, type of tutor, or number of hours on the reported effect sizes included in the study.

However, Leal, Johanson, Toth, and Huang (Leal, Johanson, Huang, & Toth, 2004) had earlier written that "tutoring is most effective when five specific factors are the foundation for tutoring: one-on-one tutoring, supervision by certified reading specialists, intensive instruction, programs that are assessment based, and regular reflective evaluation on the part of the tutor" (p. 76). As such, there continues to be general disagreement in the literature around the benefit of providing tutoring in a one-on-one setting. In contrast, researchers do agree that individual or small-group tutoring sessions should follow in support of the reading instruction given by the general education teacher rather than supplant it (Al Otaiba, Schatschneider, & Silverman, 2005).

Schools often lack the resources to hire additional staff to support classrooms and few classroom teachers have the time to provide tutoring to their students before, during, or after the school day. As a result, there has been increasing interest in evaluating tutoring models that use community members rather than certified professionals in tutoring roles (Morrow & Woo, 2001). In fact, research has shown that through

experience and a little follow-up coaching, a potential tutor can increase efficacy in supporting desired student outcomes (Tschannen-Moran & McMaster, 2009).

For example, in their 2000 meta-analysis, Elbaum and colleagues reported that 6 of the 29 studies they reviewed used community volunteers as tutors. The study reported the mean weighted effect size of studies using community volunteers as tutors was 0.59, with a range from –0.25 to 0.98. Higher effect sizes were reported among these six programs when authors also described tutors' training.

Also, Wasik (1998) reviewed 17 studies to explore the efficacy of using volunteers rather than trained and certified professionals as tutors. Volunteers in this study consisted of a mix of retired senior citizens, college students, suburban mothers, and members from the community. Some volunteers received stipends and some did not. Most of the tutors were not experienced teachers.

Three of the studies included in the Wasik (1998) analysis used randomized-controlled experimental-comparison group designs. These three programs were the Howard Street Tutoring Program (Morris, Shaw, & Perney, 1990), the Intergenerational Tutoring Program (American Academy of Arts & Sciences, 1999) and the School Volunteer Development Project (U. S. Department of Education, 1979). A fourth study using a similar design was completed of the Start Making a Reader Today (SMART) program, which is directed at K-2 students having "trouble learning the basics" (E. Baker, 2014). In fact, the SMART program included tutors serving in a much less structured environment and with much less training impacting far more students.

More recently, Ritter, Denny, Albin, Barnett, and Blankenship (2006) reviewed 21 studies that used randomized-controlled experimental-comparison group designs to investigate the effectiveness of volunteer tutoring programs. The results of this meta-analysis revealed that the average effect size of volunteer tutoring on reading outcomes was 0.30. The studies included in this review all targeted elementary school students. It is also worth noting that through secondary analysis, the authors determined that the effects did not differ by type of tutor (e.g., parents, college students, or community volunteers) or grade level of the students served.

There is still, however, concern about the potential lack of effectiveness of untrained individuals supporting students in need. In other words, do students who are assigned to ineffective tutors face a significant cost in effort with no or muted benefits? One researcher investigated the uncertainty of the effectiveness of noncertified tutors and determined that differences in outcomes can be traced to individual differences in their teaching skills (Juel, 1996). This author analyzed video and audio recordings of tutoring

sessions in which tutors (who were college students) tutored first-grade children in reading. In this case, students who showed larger reading gains received significantly more scaffolding during reading and writing experiences and explicit cognitive modeling of reading and writing from their assigned tutor (Juel, 1996).

One research team studied the impact of volunteer tutors who both effectively and ineffectively implemented a curriculum based on researcher-developed tools for afterschool tutoring (Vadasy, Jenkins, Antil, Wayne, & O'Connor, 1997). The study's post-test included a battery of a number of instruments related to achievement disseminated to students. The research team completed a multivariate analysis of covariance, controlling for a pre-test that compared the post-test results of three groups (high curriculum implementation, low curriculum implementation, and control). The study reported effect sizes for the high-implementation group of 0.89, 0.98, and 1.40 on three reading measures. They then combined the low-implementation and control groups, as they reasoned that neither of these two groups received the intervention as it was designed to be used. They did Helmert contrasts to find that students with tutors who highly implemented the curriculum had significantly higher scores than the new combined group.

An additional study investigated volunteer tutoring related to implementation adherence in tutoring sessions (Vadasy, Jenkins, & Pool, 2000). In this study, 23 first-grade students at risk for learning disabilities received the intervention while another 23 were in a control group and received no intervention. Almost all were students in a minority race/ethnic group, while about two-thirds were Title 1 students. In this study, authors observed tutoring sessions once per week, providing feedback and modeling strategies as needed, and reported an implementation adherence of 89% on six established criteria. The authors reported that providing more training before tutors began working with children along with increased supervision resulted in more accurate

implementation than they had measured in the past. The authors credited the constant supervision with positive outcomes as well as the more lengthy training and speculated that retaining tutors for more than one year could be even more successful.

The resulting impact on students was measured using pre-test and post-tests evaluating phonological skills (2000). One post-test occurred at the end of first grade and the second occurred at the end of second grade. Though the study used a composite Z of pre-test scores to adjust post-tests (because some post-tests of interest did not have that same measure administered as a pre-test), a multivariate analysis of covariance indicated significant group differences at the end of the year. In fact, tutored students significantly outperformed untutored control students on measures of reading, spelling, and decoding (effect sizes ranged from 0.42 to 1.24). However, the treatment effects diminished at follow-up at the end of second grade, although tutored students continued to significantly outperform untutored students in decoding and spelling.

The authors reported excitement about the findings as the study results fell in about the median range as compared to the effect sizes found in other studies that might have used teachers or other certified teachers as tutors (Vadasy et al., 2000). However, they also noticed that some students still fell behind and did not make that much progress, noting other studies with a similar observance (O'Connor, 1997; Vandervelden & Siegel, 1997; Vellutino et al., 1996). They noted that Juel (1996) suggested that interventions longer than one-year may be needed for many of these students.

Tutoring Curricula and Settings

Most studies of tutoring impacts have focused on the use of some type of standardized curriculum, even if it was used to support an unlicensed volunteer tutor in their work with students. In fact, in one team's meta-analysis (Wanzek et al., 2015), the authors reported that no studies examining the effects of non-standardized interventions were available to include in their synthesis. Instead the authors investigated the difference in effect between highly standardized interventions (i.e., few or no modifications to the curricula) and those with less standardization (i.e., opportunities for the tutor to respond to students' needs in the skills and strategies taught).

An example of a study of standardized curriculum was the Burns, Senesac, and Symington (2003), evaluation of the effectiveness of the Helping One Student to Succeed (HOSTS) program. The HOSTS program uses one-on-one tutoring (termed mentoring by program authors) and individual lesson plans (developed by a HOSTS-

trained teacher using a HOSTS-designed database and designed to address the student's developmental levels, learning styles, and the specific learning objectives to be achieved). The authors found that at-risk students in the treatment group outperformed those in the control group with a mean effect size of 0.26 across reading measures. In addition, the authors later documented longer-term positive effects through a two-year follow-up study (Burns, Senesac, & Silberglitt, 2008).

Another example of such a program is the Book Buddies model, which used older adult volunteers as tutors for first graders in a high-poverty urban setting (Meier & Invernizzi, 2001). The Book Buddies program involves coordinators who are paid and certified professionals who supervise each session and provide ongoing training to volunteers. In addition, the Book Buddies coordinator assesses each student individually, writes individualized lesson plans for each one, and gathers the appropriate materials for each lesson. In effect, though lessons are personalized, the tutoring framework itself was fairly standardized. The authors reported that the children who received 40 Book Buddies lessons were significantly stronger in reading skills than the comparison group at the conclusion of tutoring, but that effect disappeared by the end of the school year a few months later.

Meier and Invernizzi (2001) also reported that their results underscored the following: "(a) the use of research-based instructional elements to produce significant progress in reading achievement; (b) the need for a well-structured lesson plan in which the content and delivery of instruction are carefully planned; (c) the need for intensive, ongoing training and supervision of tutoring under the guidance of a knowledgeable reading specialist; (d) the need for frequent and regular tutoring sessions; and (e) the need for careful evaluation, assessment, monitoring and reinforcement of progress."

The study of Project MORE (Osborn et al., 2007) specifically investigated the implementation of two individualized reading interventions named HOSTS and Reading-Tutors, both of which had a standardized tutoring framework. Both were one-on-one and were reported by programmers to be comprehensive in supplementing the regular reading program. Mentors in this study were volunteers from the community as well. The HOSTS sessions were split into three 10-minute sessions: guided reading instruction, spelling and vocabulary, and specific skill development.

The Reading-Tutors curriculum was an online repository that included 400 comprehensive lesson plans and assessment tools patterned after DIBELS (Osborn et al., 2007). Overall, both Reading-Tutors and HOSTS students had statistically significant month-for-month reading gains on two instruments when compared to students not

receiving an intervention. The authors reported that this study demonstrated that a cohesive remedial intervention that delivered the Reading-Tutors and HOSTS programs using volunteers supported by school staff could lead to a statistically significant increase in the reading achievement level of students, even those with serious reading disabilities. It should be noted, however, that the control group in this study consisted of students in different schools receiving a different general education curriculum than the intervention students.

In addition to investigation into the use of specific curricula, the literature provides programmers with some guidance regarding the most appropriate settings for student tutoring. This guidance includes a discussion of one-on-one versus small-group settings, as well as discussions about location of tutoring in the building and conducting sessions during or after school. In this review, we will investigate the findings around outcomes in small-group versus one-to-one settings.

It is worth noting that the U.S. Department of Education's Institute of Education Sciences reports that research evidence clearly supports "one-on-one tutoring by qualified tutors for at-risk readers in grades 1-3" (Institute of Education Sciences, 2003, p. iii). However, providing one-on-one tutoring to children has a significant resource cost in most buildings and programs.

Vaughn, et al. (2003) conducted a comparative study of 77 students taught in teacher-student ratios of 1:1, 1:3, or 1:10. The study compared pre- to post- achievement scores across a school year. Findings from this study indicated that, overall, students made significant gains following the intervention across a number of skills, though gains were greater with 1:1 tutoring than 1:10. However, 1:1 tutoring was not superior to 1:3 tutoring for any outcome measure. Effect sizes ranged from 0.004 to 0.20 between the groups depending on the skill measured. Using an additional post-test, the authors also confirmed that the effects of the intervention remained five weeks after the intervention ended.

Another team (Cawthon & Maddox Jr, 2009) studied the impact of one-on-one versus small group tutoring for reading (and in this case writing). This study is unique among those reviewed so far for a number of reasons: 1) tutors were certified educational therapists, 2) sessions were 80 minutes and offered twice a week, 3) students were enrolled at private, not public, schools. Despite these differences from other studies, this study's investigation between group settings is enlightening.

The authors studied five unique outcomes as measured by different instruments, four of which were related to reading and one of which was related to writing (Cawthon &

Maddox Jr, 2009). Of the four reading-related outcomes, one-on-one tutoring effect sizes were superior to those of small-group tutoring for three of the four skills measured. However, the one-on-one tutoring group had inferior outcomes on the word recognition measure compared to those of small-group tutoring. It should be noted that effect sizes were between 0.26 and 0.68 for small-group instruction and were typically within 0.20 of one-on-one effect sizes, even when less effective.

We mentioned the evaluation of Dorn and Allen (1995), who studied the implementation of a small-group tutoring initiative called Reading Recovery. That study found that small-group instruction was beneficial for children who needed less supplemental help if those programs generally extended over a longer time period than one-on-one sessions.

Harrison (2002) also studied the implementation of Reading Recovery and a specially designed small-group model. Her conclusions were similar to those of Dorn and Allen: (a) the lowest-achieving children need a one-on-one setting, (b) small-group instruction is more beneficial for children needing less supplemental help, and (c) children served in small groups tend to need longer interventions, usually for the full school year.

Also mentioned earlier, Elbaum et al. (2000) found one-on-one instruction as a supplement to classroom teaching to be "the most effective way of increasing students' achievement" (p. 605) and noted that it has been "validated by empirical research, especially for students who are considered at risk for school failure or have been identified as having reading or learning disabilities" (p. 605).

In addition, of the previously mentioned studies that showed significant positive academic change for tutored students as compared to students not receiving tutoring, four investigated one-on-one tutoring. These included the Vadasy and colleagues (1997), Allor & McCathren (2004), Osborn, and colleagues (2007), and Lee, Y. S., Morrow-Howell, N., Jonson-Reid, M., & McCrary, S. (2010) studies. All showed significant and positive effects of one-on-one tutoring.

In a study specific to the tutoring support of Kindergarten students, the Tutor Assisted Intensive Learning Strategies (TAILS) program evaluators (Al Otaiba et al., 2005) investigated achievement impact. In this study, the tutoring lasted all school year, and though not specified, was likely delivered in a one-to-one setting given the ratio of tutors to students.

The authors reported that tutors used an unused classroom or, in one school, the lunch room before and after lunch time (Al Otaiba et al., 2005). Each TAILS session lasted 30

minutes. The study design consisted of students at four schools and random assignment to one of three conditions: (a) tutoring four days a week, (b) tutoring two days a week, or (c) a control condition that provided small-group storybook reading two days a week. It should also be noted given previous conversations that the tutors in this study received three hours of training over three sessions (two before initiating the intervention and one the week after).

Children were administered reading and phonemic awareness pre- and post-treatment tests (Al Otaiba et al., 2005). To investigate the effects of the two experimental tutoring conditions, the authors analyzed the pre-test–post-test means from the three randomized groups with a series of 3 × 2 analyses of variance. Analyses revealed that students in the four-day condition outperformed students in either the two-day or control conditions on the three reading measures used, with effect sizes ranging from 0.79 to 0.90. Additionally, the authors reported that both the two-day and four-day students outperformed students on "Blending Sounds" by a significant margin (ES = 0.68).

The program effects were achieved despite the limitations mentioned by the authors (Al Otaiba et al., 2005). They reported that there was no funding to hire a reading specialist who could tailor lessons for individual students (a key part of other programs) and that schools were challenged to protect the time for tutoring beyond the 90-minute literacy block and to devote a designated quiet area for tutoring.

Effectively Selecting, Training, Supervising, and Supporting Tutors

As a result of the findings reported above, researchers have focused on the opportunities available to select, train, supervise, and support uncertified volunteer or non-professional tutors in specific skills and strategies. At the same time, and as a result of limited resources, most programs appeared to consist of just a small amount of training or follow-up coaching following tutor selection. Supervisory settings were highly varied. Below, we will explore studies that targeted these variables.

Specific to AmeriCorps reading programs, a team evaluated the Minnesota Reading Corps (Markovitz, Hernandez, Hedberg, & Silberglitt, 2014), the largest state program in the country. The program recruits, trains, places, and monitors members as they implement tutoring interventions for students in kindergarten through third grade. The program supports a paid position at each school to supervise its local Corps of members. Unlike most programs previously mentioned, the program fits snugly into an existing school framework. In this case, the tutors provided one-on-one tutoring as Tier 2

interventions, or targeted group interventions for students who fell below the expected levels of accomplishment in the three-tier response-to-intervention framework (Johnson & Pesky, n.d.).

The research team (Markovitz et al., 2014) initiated a randomized-controlled experimental-comparison group design with 1,300 students at 23 schools. The authors found that Kindergarten, first-grade, and third-grade students who received the intervention showed significant achievement gains as compared to students in control groups. These gains were steady across different racial groups and socio-economic statuses. Effect sizes ranged from 0.1 (third-grade) to 1.1 (Kindergarten) across grade levels.

A study investigated the Experience Corps® (EC), a program that places older adult volunteers with 15-32 hours of training in elementary schools to tutor students who are struggling readers (Y. S. Lee et al., 2010). In the case of EC, there were 2,000 older volunteers nationwide serving 20,000 students in 23 cities at the time of the study. There is a curriculum, however, it is chosen locally. For example, Book Buddies, Reading Coaches, and the Brigance Inventory of Basic Skills were used in three different cities.

Tutors spent two to four sessions a week with the students for 30-40 minutes per session, averaging between 35 and 58 sessions for each of the three locations included in the study (Y. S. Lee et al., 2010). Students were identified for tutoring based on the professional judgment of educators in their school and tutoring was provided in one-on-one environments.

The authors implemented a randomized-controlled experimental-comparison group design involving 883 students in the three urban cities (Y. S. Lee et al., 2010). Students were assessed with standardized reading measures at the beginning and the end of the school year. In addition, teachers were also asked to complete two surveys about their students.

The authors analyzed program effects using Generalized Estimating Equations to adjust for clustering effects (Y. S. Lee et al., 2010). Findings indicated that intervention students made statistically greater gains over the academic year on passage comprehension and grade-specific reading skills, especially those who received 35 tutoring sessions or more. The authors reported effect sizes between 0.10 and 0.17 on the various measures.

Mentioned previously in this review, Vadasy (2000) also developed a structured program for volunteer tutoring. That program used tutors who were paid $5 an hour,

who had received 14 hours of training, and who used one-on-one reading tutorials for low-achieving first grade students. The program was led by non-professional tutors providing one-on-one tutoring for 30 minutes, 4 days a week, for an entire school year. The authors developed 100 scripted lessons that targeted decoding, rime analysis, and story reading. Students were tested every 10 lessons to ensure mastery. Project staff administered these assessments. The tutoring intervention studied lasted for 27 weeks, with children attending between 54 and 89 sessions.

Tutors maintained daily logs of attendance and lesson progress. At the end of the year, tutored students significantly outperformed untutored control students on measures of reading, spelling, and decoding. The authors reported effect sizes ranging from 0.42 to 1.24 (Vadasy et al., 2000).

Vadasy's team also implemented tutoring with volunteers who were initially trained for six hours, followed by three hours of training once they had experience in the field (1997). The research team developed a curriculum that incorporated research-based components that promoted phonological awareness. The program was developed to put untrained tutors with students in one-on-one environments. Specifically, they designed highly structured lessons with explicit scripts for tutors to follow in teaching each skill, developing 100 30-minute lessons that included as many tasks as possible that had been identified in previous studies to contribute to gains.

The tutors who implemented these curricula were parents, grandparents, college students, and high school students (Vadasy et al., 1997). Tutoring began in November and took place after school. The intervention ended in May at a maximum of 23 weeks and 53 hours of instruction.

The authors compiled a Z score of all screening and pre-test measures and then averaged them to make a composite (Vadasy et al., 1997). This composite was used in a multivariate analysis of covariance comparing treatment and control post-tests. Tutored students' post-test means exceeded those for the control group on all measures; effect sizes ranged from 0.15 to 0.56.

Improved Student Attitudes toward Reading

In support of continued tutoring-related academic improvement, it is desired that students also have changed attitudes toward reading as a result of tutoring. This outcome is important, as the research literature has documented a link between

aspects of attitude, like motivation, to greater achievement (Elley, 1992; Gambrell, Palmer, Codling, & Mazzoni, 1996).

The McKenna (1994) model proposes that three principal factors influence attitude change: (a) beliefs about the outcomes, (b) beliefs about the expectations of others in light of one's motivation to conform to those expectations, and (c) the outcomes of specific incidents of reading. Practically speaking, this model suggests that students have positive and negative reading experiences and develop beliefs about their own abilities and about the act of reading itself. The implication over time is that these beliefs permanently inform the student's attitude towards reading.

Research has shown that it is certainly not assured that classroom teachers will positively influence student attitudes towards reading during the traditional in-classroom reading block. In fact, Schmitt (2009) investigated the impact of "teacher attitudes" on student attitudes towards reading and found a significant correlation, negative and positive. Unfortunately, not all classroom teachers portray positive attitudes towards reading in their professional demeanor. In another study, Applegate and Applegate (2004) surveyed 379 students and found that 51.5% were considered unenthusiastic readers by their standards. "It was clear that significant numbers of respondents were affected, either positively or negatively, by the instruction they received during their early school years."

Unfortunately, research has also established a trend between the reading attitudes of good readers and poor readers, a gap that tends to widen with age (McKenna, 2001), including a gradual worsening of attitudes from 6th grade to 8th grade (McKenna, Conradi, Lawrence, Jang, & Meyer, 2012). For this reason, it is imperative that student attitudes be supported in the early elementary years to elevate a student's attitude to a level that can sustain the future drop.

Children's attitudes about reading have been studied from several different approaches, but typically in reaction to whole-class interventions. For example, research has explored children's attitudes about literacy in relation to specific literacy interventions. For instance, McKenna, Stratton, et al. (1995) explored whole-language and basal reading instruction and their impact on student attitudes.

Even so, how do research teams and evaluators determine student attitudes about reading? In fact, researchers have developed a number of different instruments to effectively measure student attitudes. For example, the Reader Self-Perception Scale (RSPS) is an assessment tool developed by Henk and Melnick (1992) that focuses on the psychological components of readers' perceptions of their own reading abilities.

Gambrell, et. al., (1996) extended the work of Henk and Melnick by focusing on readers' self- concept and task value in a tool called Motivation to Read Profile.

The Motivation to Read Questionnaire (MRQ) is a literacy assessment developed by Baker and Wigfield (1999) and focuses on engagement and motivation to read. The three larger areas within the MRQ are competence and efficacy beliefs, goals for reading, and social purposes for reading. These three areas were assessed with 50 total items and analysis was presented within 11 categories. It should be noted that Unrau and Schlackman (2006) administered the MRQ only with older children
(6th-8th graders).

Many studies utilized another tool that assessed the frequency and scope of students' reading habits. The Reading Activities Inventory (RAI) developed by Guthrie, McGough, and Wigfield (1994) was designated primarily to measure the amount and breadth of students reading. As a result of slightly differing foci of study, this tool investigates reading of specific types of media and is less strong in measuring student attitudes about reading.

Finally, McKenna (whose model of attitude change is mentioned at the beginning of this section) and Kear (1990) developed a public-domain instrument, the Elementary Reading Attitude Survey (ERAS), for use by researchers, teachers and school administrators alike. The authors, and many other researcher teams (Applegate & Applegate, 2004; Navarra, 2011; Parker, 2004; Roth, Worrell, & Gabelko, 2002; Seitz, 2010; Worrell, Roth, & Gabelko, 2006) have used the ERAS with young children to determine the overall attitude levels of samples or classes and to provide insights into the reading habits and achievement levels of individual children. A number of authors (Henk & Melnick, 1992; Kazelskis et al., 2005; Kazelskis, Thames, & Reeves, 2004; McKenna et al., 2012; McKenna & Kear, 1990) have provided evidence that the ERAS exhibits solid validity and reliability characteristics. For example, in one study of more than 18,000 students, internal consistency coefficients for the recreational reading subscale ranged from 0.74 to 0.89. Adequate construct validity was reported by relations to library card possession and hours of reported television watching (McKenna & Kear, 1990).

To date, there has been little study of the impact or effect of tutoring interventions on student attitudes towards reading. In fact, none could be uncovered that specifically investigated a tutoring intervention provided by an adult. The closest match was an evaluation of an "intergenerational reading aloud program," which had a positive effect on elementary student's reading attitudes, motivation and achievement (Pena, 2008).

One study attempted to determine if peer tutoring, rather than adult-to-student tutoring, made an impact on attitudes. In this case, cross-age peer tutoring of fifth graders with learning disabilities and Kindergarten students would have an effect on reading attitudes (Davenport, Arnold, & Lassmann, 2004). Results indicated that the program had a positive effect on reading attitudes of the Kindergarten students. Another study, also regarding peer tutoring, actually turned its lens of focus on the peer doing the tutoring, rather than the student receiving the tutoring (Hamman, n.d.). In this study, a very small number of students (n=8) were found to have benefited around "academic efficacy," or a belief that one can succeed in the classroom, from providing tutoring.

The purpose of the present study is to further investigate the impact on K-3 elementary students of volunteer adult tutoring in one-on-one and small-group settings across the state of Mississippi. Specifically, two research questions are asked. First, what was the America Reads - Mississippi program's impact on improving student academic gains in reading during the 2015-2016 school year? Second, what was the impact of the same program on student attitudes towards academic and recreational reading?

With limited resources, building administrators are often unable to provide tutoring interventions to all struggling readers at all grade levels. The findings of this study could provide information on the feasibility of administering a state-wide program utilizing AmeriCorps volunteers as tutors in the subject area of reading. Policy makers may refer to the findings as decisions are made regarding distribution of resources aimed at interventions for students struggling in reading. This information could be critical for policy makers and others who want to develop their own state-based or regional reading tutoring programs.

METHODS

Schools and Settings

The study involved students at 20 schools in the state of Mississippi. All schools were returning ARM partners with at least a two-year history with the program. Some of these schools utilized members at four grades involved in the study, while some schools concentrated members in one grade or chose to utilize members in two to three grades. Even in locations where tutoring was spread across grades, members served in just one grade. We share these table in Table 1.

Table 1. Grade distribution of evaluation students across schools

School	K	1	2	3	Total N
Amite County Elementary	19	16	15	26	76
Coffeeville Elementary	8	10	23	17	58
Eva Gordon Elementary			11	21	32
Fair Elementary	59	31	34		124
Franklin Lower Elementary	15	15	18	21	69
Goodman-Pickens Elementary				19	19
Hawkins Elementary	9	11			20
Holly Springs Primary		31	10	6	47
Jefferson County Elementary	18	25	18	24	85
Lumberton Elementary	6	9	17	17	49
Maddox Elementary	27		22		49
North Pontotoc Elementary		26	49	22	97
Oak Park Elementary			34	10	44
Okolona Elementary	16	18	27	6	67
Osyka Elementary		3	8	4	15
Petal Elementary		83	49		132
S.V. Marshall Elementary		15	8	16	39
Sherman Elementary			32		32
South Pontotoc Elementary		38	15	14	67
Warren Central Intermediate				20	20
Total N	**177**	**331**	**390**	**243**	**1141**

The schools contributed a maximum of 132 students and a minimum of 15 students to the impact evaluation.

The schools selected for inclusion in the impact evaluation were located across all regions of the state. These schools had variable pass rates on the 3rd grade reading assessment—the MKAS state test. Looking at a more all-encompassing measure of school success in supporting positive student academic outcomes, these schools were also variable in their most recent state report card grade for reading instruction. The highest rating achieved by a school was a B, achieved by two schools. Schools were more likely to have received a C (5 schools) or a D (9 schools) with four schools receiving the lowest grade, an F, from state auditors. These data are shown in Table 2.

Table 2. State Report Card Grade For Reading Instruction across schools

School	2014 2015 Reading Grade
Amite County Elementary	D
Coffeeville Elementary	D
Eva Gordon Elementary	F
Fair Elementary	D
Franklin Lower Elementary	F
Goodman-Pickens Elementary	F
Hawkins Elementary	C
Holly Springs Primary	D
Jefferson County Elementary	D
Lumberton Elementary	D
Maddox Elementary	D
North Pontotoc Elementary	B
Oak Park Elementary	D
Okolona Elementary	B
Osyka Elementary	D
Petal Elementary	C
S.V. Marshall Elementary	F
Sherman Elementary	C
South Pontotoc Elementary	C
Warren Central Intermediate	C

Students

A grand total of 1,141 students were included in our study sample and either received the tutoring intervention or served in a control or comparison group. As mentioned previously, the students were enrolled in Kindergarten, 1st grade, 2nd grade, and 3rd grade across the twenty schools (not all schools served students at all grades). These students became eligible to receive tutoring, and thus eligible for the impact evaluation, by scoring between the 10th and 45th percentile on their initial screening assessment

and by returning a signed parental/guardian consent form highlighting the tutoring available as well as the study.

Of the students who scored between the 10th and 45th percentile, returned the consent forms, and were selected as a part of the evaluation sample, 49 percent were female and 51 percent were male.

Students with disabilities were not accepted into the program. It is ARM policy that its members do not serve these students, and that these students instead be served by professionals with training and experience related to academic outcomes and this population. Along these lines, students in third grade who had not passed the previous year's 3rd grade assessment at of the start of the school year were also not included in the study sample. Again, ARM policy specified that these students receive support from professionals with training and experience related to preparing these students to pass a mid-fall assessment retake in order to advance them into 4th grade.

We did not receive information regarding ELL status and FRL status from all locations, however, for the sample from which we did receive this information, we found that just 5% of students were reported to be English language learners and 96% of students were reported to receive free or reduced price lunch. It is likely that the mean rate for free or reduced price lunch is even higher, as we did not receive these details from a number of schools that are no longer asked by the state to collect such information as they serve neighborhoods with overwhelming poverty.

The student racial/ethnic demographic information indicated that as a whole, these students were most likely to be African American (66%), followed by White (29%), Hispanic or Latino (4%), or Native American or Asian (<1%).

Table 3. Student demographic makeup across evaluation sample

Variable	Rate
Female	49%
Male	51%
ELL Status	5%
FRL Status	96%
African-American	66%
White	29%
Hispanic or Latino	4%
Native American	<1%
Asian-American	<1%

Program Design

Established in 1998, ARM is a national service AmeriCorps program that addresses the issue area of education. ARM is funded through a grant from the Corporation for National and Community Service (CNCS), matching funds provided by partnering school districts, and funds provided through the Mississippi legislature as appropriated to the program's legal applicant, the Mississippi Board of Trustees of State Institutions of Higher Learning (IHL).

ARM is dedicated to improving the reading skills of students, encouraging public awareness and support of literacy, and helping to increase the number of certified teachers in Mississippi. ARM AmeriCorps members strive daily to achieve this mission and meet the Program Performance Measures. The primary performance measure is related to school-day reading tutoring. ARM members tutor primarily K – 3rd grade students (one-on-one and in small groups) in reading during school and in extended-day/year programs.

ARM members provide tutoring in reading to students with the greatest academic need. ARM members meet with the classroom teacher and reading specialists to implement school-selected, research-based interventions and to regularly assess student gains and/or additional needs. ARM AmeriCorps members also recruit volunteers to assist with reading activities in the classroom, service projects, and service-learning activities; help to strengthen the home-school connection; plan and implement national service day projects; and participate in member training and development opportunities provided by both the school service site and the ARM program.

Administration

ARM is administered statewide with oversight from the Volunteer Mississippi and the IHL Office of Academic and Student Affairs, and is subject to the policies and procedures of each of these entities. AmeriCorps members receive program support from one of three regional university offices located at Jackson State University, Mississippi State University, and the University of Southern Mississippi. Overall state management is administered from the ARM state office located in Jackson, Mississippi at the IHL offices.

These regional offices, with support from the state office, assist schools in the selection of members, train these members, monitor their progress during the course of the school year, and conduct quarterly observations of implementation adherence.

Tutor recruitment

Members are recruited, interviewed, and selected by the local school sites themselves, in a manner comparable to the method used when hiring traditional staff. Members must be United States citizens, a national, or a legal permanent resident, be at least 17 years of age, and have a high school diploma or an equivalency certificate in order to enroll as an AmeriCorps member.

Additional relevant minimum requirements include proof that any potential member:

a) has passed the Reading section of the ACT WorkKeys Test or has completed two years of college in order to comply with AmeriCorps regulations for tutoring programs.

b) understands that, as required by the Corporation, America Reads – Mississippi (through its regional offices and partnering agencies) will conduct the following criminal history checks, with which the applicant/member must comply and MUST PASS as a condition of his/her service with zero reportable incidents:

The member's term of service is made up of 11 months of service between August 1 and June 30. If a member has served in ARM for a full-time one-year term (or in another AmeriCorps program for a year), they can be eligible for service in ARM as a second-year AmeriCorps member if they successfully completed the first year of service. The only possible but not automatic exceptions to this are: (1) if a member did not complete the required service hours due to approved personal compelling circumstances, or (2) if the member BOTH served within 95% of completion AND receives a recommendation from the current service site AND AmeriCorps program staff. These exceptions will not extend beyond one service term.

The member also understands that they will be evaluated at mid-year on the quality of service provided, as well as other aspects of the AmeriCorps member position.
Some of these aspects include:

- *punctuality and regular attendance at school service site*
- *attending program-sponsored trainings and events*
- *presenting a positive attitude while working with program staff, school staff, and students*
- *timely, accurate, and complete submission of timesheets and program reports, and*
- *limited disciplinary action due to policy infractions*

Training

ARM members attended program-required and various school-sponsored trainings throughout the service year. The program desires to equip members with programmatic information and foundational reading instruction skills that members can build upon over the course of the program year. Ongoing school-level training is provided by members' service sites on a consistent or as-needed basis.

All members were required to attend a one-day regional orientation where they learned more about the America Reads - Mississippi program, received training on completing time sheets and documenting tutoring time, learned about community service requirements, and participated in team-building activities with fellow members. The members also attended an additional program-sponsored tutor training facilitated by Mississippi Department of Education (MDE) Regional Literacy Coordinators. At these trainings members received foundational information on the five components of reading instruction (phonological awareness, phonics, vocabulary, fluency, and comprehension) as well as strategies and resources for planning tutoring sessions to address these components. Tutoring supplies and instructional materials were distributed to all members at this training.

Members were required to attend school-sponsored instructional meetings and professional development throughout the program year to learn more about instructional strategies and resources that can be utilized for tutoring. Some members reported attending additional training sessions facilitated by MDE presenters as part of their school-sponsored training. These members received tutoring kits with literacy resources and used them in their tutoring sessions.

Prospective team leaders at each school site were also invited to attend a leadership training hosted by the program. This training enabled members to expand their leadership capabilities as they listened to a successful local business owner as he discussed leadership principles and shared his experiences. Members in attendance also heard from other educational professionals who discussed the importance of goal-setting and positively impacting students.

In February, all members attended regional trainings and were trained on creating customized intervention plans for their students and on administering the Elementary Reading Attitude Survey. Members received charts that listed their students' baseline and mid-year assessment scores as well as target end-of-year scores. Members also received a training refresher on submitting electronic reflection logs to include students' tutoring time and reviewed the remaining program requirements for the service year.

At the conclusion of the service year, members attended regional trainings that focused on student achievement, life after AmeriCorps, accessing the CNCS education award, and setting future goals. Additionally, members were recognized for their service at these trainings and congratulated by school site supervisors, community partners, and parents in attendance.

Supervision

In addition to other responsibilities, these regional entities train one or more site supervisors at each location to assist with the day-to-day management of members at the local level. These supervisors provide their time "in-kind" to the program and hold a variety of roles within the school, including principal, assistant principal, lead teacher, and academic coach. In implementation, members receive a varied level of support from these individuals depending on location, ranging from support only in administrative tasks such as logging time sheets, to support with professional skills including pedagogical knowledge or content area coaching.

ARM members are AmeriCorps participants and are protected from many tasks associated with employment. As such, members are not allowed to perform an employee's duties or otherwise displace employees. In this case, members in schools are not meant to serve as classroom leads, or even assistant teachers, and should have their mission as literacy "tutors" protected and respected by local entities.

Record keeping

All members communicated to central administrators information about their tutoring, including the details regarding the students served, the amount of time spent in session, and the exact setting (one-on-one or small group) that they served in. They also communicated to regional administrators a more granular level of information about their service as kept in a service binder. These details included specific interventions used and notes about progress. As a result of this communication, regional and central administrators were able to intervene as needed in instances where members were not able to access students for meaningful tutoring (e.g., the student was not made available by the classroom teacher, or the students were sent in numbers too large for small group practice).

Intervention

Members served students with one-on-one and small-group tutoring that occurred either outside of the classroom or in a space set aside within the classroom in a

distraction-free area. Members were assigned the responsibility of tutoring each student for 30 minutes a day at least 4 days a week.

Site supervisors at the local level who are employees of the district and who are familiar with both the students and members were responsible for the assignment of students to members. Often, members were assigned all students eligible for tutoring within a given teacher's classroom, however some students were assigned directly to members even though they sat in different classrooms.

Members did not use a standardized curriculum or strategy within sessions but instead were charged with the responsibility of collaborating with classroom teachers to develop lessons to directly support the student's learning during traditional reading block instruction.

Treatment fidelity

As mentioned previously, central administrators and regional administrators were able to intervene following member reports of issues accessing students in a meaningful environment for tutoring in a one-on-one and small-group setting. However, administrators also completed quarterly observations of the tutoring environment within each building. These administrators completed an implementation adherence rubric on each of their visits during the late winter and early spring of 2016. This tool asked the rater to assign between 1 and 5 points to each building in seven key areas of implementation. As one might expect, the average adherence score ranged among buildings by 16 total points. However, only four buildings averaged less than an "average" rating across the assessment categories. Two buildings averaged more than an "above average" rating across all areas. We present these data in Table 4.

We also asked members to self-report on the quality of the setting that was developed for each of them within their school building. These items mapped back to our observation categories in an effort to triangulate findings. When member responses were aggregated at the school level, we were able to add an important data point related to the quality of the setting that the member served in, beyond what an administrator might have seen on one, two, or three observation visits.

As shown in Table 4, we found the values reported by members to be moderately correlated with the values reported by administrators at a Pearson Product-Moment Correlation Coefficient of 0.40 value. This coefficient was significant at the .10 level for a two-tailed test. The coefficient was not significant at the .05 level for a two-tailed test.

Table 4. Treatment fidelity as measured by member report, and direct observation

School	Member Report	Observation Adherence Score
Amite County Elementary	0.5	25.3
Coffeeville Elementary	1.5	26.0
Eva Gordon Elementary	1.5	27.0
Fair Elementary	0.8	25.0
Franklin Lower Elementary	1.1	23.0
Goodman-Pickens Elementary	0.9	24.0
Hawkins Elementary	1.2	20.5
Holly Springs Primary	0.5	26.0
Jefferson County Elementary	1.8	24.3
Lumberton Elementary	1.0	18.0
Maddox Elementary	1.4	30.0
North Pontotoc Elementary	1.4	26.3
Oak Park Elementary	0.3	19.3
Okolona Elementary	1.3	21.0
Osyka Elementary	1.5	34.0
Petal Elementary	0.8	21.2
S.V. Marshall Elementary	1.6	22.0
Sherman Elementary	1.7	26.0
South Pontotoc Elementary	1.3	25.0
Warren Central Intermediate	0.8	23.0
Mean	**1.1**	**24.1**

Dosage

As a result of issues in scheduling members and students in some locations and variability in the final testing dates within buildings, the total number of weeks spent tutoring students differed across buildings, and even within buildings. All students were tutored for a minimum of 10 weeks, and the mean duration of the intervention was 22.6 weeks. The maximum number of weeks a student spent receiving tutoring prior to final testing was 29 weeks.

As one might expect, this also caused variability in the total number of hours that each student spent receiving tutoring. In fact, the amount of time students spent in tutoring during the school year ranged from just under 9 hours to 110 hours. The mean duration of tutoring time across all students was just under 48 hours across the school year.

We also know that the variation in member ability would likely impact changes in student outcomes that could be attributed to tutoring implementation. For this reason, we analyzed ARM administrator mid-year and end-of-year member evaluations. We assessed the ARM scale for validity and reliability and then completed a principal component analysis followed by correlation examination.

Principal component analyses with no rotation, with orthogonal rotation, and with oblique rotation all resulted in the identification of three factors. It was clear that just two of these factors were related to changes in student outcomes, while others focused on outcomes at other levels or administrative tasks (e.g., participation in community service projects).

The result was a twice-administered, five-question, and four-point scale (maximum score of 20 points) and one-question and four-point scale that consisted of ARM administrator evaluative impressions of member ability. As expected, there was a fair amount of variability in ratings, from a minimum mean score of 10.2 points on the multi-question scale for one member to a maximum of 20 points (a perfect score) for four members. The average mean score across both the mid-year and end-of-year evaluation of member ability was 17.2 points. Three member evaluation scores fell more than two standard deviations (S.D. = 2.3 points) below the mean.

Control Condition

Students assigned to the control condition were not assigned to a tutor and did not receive support from an ARM member during the school day throughout the length of the intervention. However, these students did continue to receive the same general education instruction in the same classroom as students who were assigned a tutor. These non-intervention students may have received additional support from district personnel as determined at a local level.

Study Design

The evaluators implemented a three-pronged approach to the evaluation of two major outcomes during the 2015-2016 school year: 1) student attitudes towards reading, and 2) student reading/literacy achievement.

An experimental or quasi-experimental design was not possible for our analysis of student attitudes towards reading as we did not have regular access to students who did not receive the intervention. For the achievement analysis, as a result of member

assignment patterns to specific locations and the wishes of locations to have members serving at specific grade levels, we chose an experimental approach for Kindergarten and 1st grade and a quasi-experimental approach for 2nd and 3rd grades.

The experimental condition for Kindergarten and 1st grade consisted of randomized assignment of all eligible students to either the intervention or a control condition. The quasi-experimental approach for 2nd and 3rd grades involved the matching of all students on both pre-test scale scores and student growth percentile (to control for point-in-time performance). To match students, we completed calculations using Propensity Score Methods (Rosenbaum & Rubin, 1983) to reduce or eliminate the effects of confounding data. PSM are used frequently in education.

All schools were required to assign all students selected for the intervention. Next, when additional tutoring slots were available, students from the matched group could be assigned to a member. Finally, all students receiving the intervention were assigned to one of two groups: those who initiated tutoring early in the fall and those who initiated tutoring late in the fall.

In both cases, and as mentioned earlier, schools made the decision on which members to whom to assign students. Evaluator interference in normal school practices stopped at assignment of students to evaluation groups.

In addition, as a result of a lack of access to non-intervention students in our study of student attitudes, we used a post-hoc clustering of students whose outcome measures we had received. These groups were designed to explore the impact of the tutoring intervention on the student attitudes.

Pre-Test and Consent

All students completed an achievement test during the early fall (August, September) of the school year. Students were tested using the STAR battery of assessments for literacy as set forth by the Mississippi Department of Education. These measures were selected as all buildings were required to assess all students using the tools at least three times during the school year, including once in the fall for screening purposes. In addition, some of the buildings also used the assessment for progress monitoring purposes.

As mentioned previously, parental consent was a program requirement before students were eligible for assignment to an ARM member. As a result, parental consent was also a requirement for inclusion in the study. Parental consent was acquired following pre-test distribution, but before sampling. The evaluators added language to the consent

form regarding the impact evaluation and what it meant for students during the 2015-2016 school year.

Sampling

It should be noted that because of the starkly different constructs investigated by each assessment, our analysis considered them to be completely different samples of students. As a result, three distinct impact evaluation samples were created following testing among students who met study criteria (including demonstrating parent consent): 1) all students in Kindergarten and 1st grade tested on the STAR Early Literacy examination, 2) all students in 1st grade tested on the STAR Reading examination, and 3) all students in 2nd and 3rd grades tested on the STAR Reading examination. We provide a visual describing these samples and study design below in Table 5.

Table 5. Sampling procedures for each evaluation of student achievement outcomes

Sample Name	STAR Early Literacy	1st STAR Reading	2nd 3rd STAR Reading
Grades	K,1	1	2,3
Design	Experimental	Experimental	Quasi-Experimental
Sampling Method	Random	Random	PSM & School Choice
Non-Intervention Group	Control	Control	Comparison
Tutored Group(s)	Intervention	Intervention	Late Intervention, Early Intervention
Total N Baseline	231 (113 control)	277 (136 control)	633 (134 comparison)

Analysis Decision Rules

Students were removed from analysis if they were not able to complete their school year tutoring due to either student-level or member-level issues. For example, students may have withdrawn from tutoring due to parent choice during the middle of the school year. At the same time, a student's member/tutor may have exited from the program, leaving the student without a tutor during the middle of the school year.

We also removed a small number of students due to logistical issues at the school level. These included students who either were not tested at the end of the school year or were tested too early for the assessment to be considered an end-of-year examination. In addition, a number of students did not sit for the student attitudes outcome measure (a survey) in either February or March of 2016. We also removed students who did not meet a minimum threshold of tutored weeks due to school-based scheduling and testing decisions.

Member-level factors

The primary reason for student removal from analysis for member-level reasons were the physical removal of the member from the tutoring location due to a member decision to end their volunteer service or an administrative decision to exit them from the program. At the same time, there were a small number of instances where a member was unable to access students for extended periods of time due to their own lack of attendance because of illness or other reasons. These students remained in the evaluative sample unless they did not meet minimum intervention time requirements.

Student-level factors

There were a number of instances where a student may have been removed from analysis for student-level reasons. These included when a student moved or was unwilling or unable to consistently attend tutoring sessions in order to meet a minimum threshold for hours of intervention. We also did not include a small number (n=4) of students in the analysis whose change in test score was more than three standard deviations outside the mean, indicating either an incomplete test form or a lack of fair effort or guessing on one of the test disseminations.

Effect on Analysis Sample. For the analysis of student attitudes, we also did not include survey responses with more than 5 missing item answers in our analysis. In some cases, these students had an additional form that could be used in analysis, and in other cases, the students were removed from the analysis sample. A total of 302 students did not complete an eligible survey, 43 of them students who received an intervention. When students completed more than one survey (possible but not frequent), their responses were averaged across both forms to further strengthen reliability.

The total number of students removed from the analysis of student outcomes due to either student or member withdrawal from the intervention or the evaluation equaled 74 before the spring survey distribution and 137 prior to the end-of-year STAR assessment. A total of 32 additional students were excluded from the achievement analysis for the reasons explained above, while 4 students were excluded from the survey analysis.

Table 6. Details regarding preservation of evaluation samples across the school year

Sample Name	STAR Early Literacy[1]	1st STAR Reading[1]	2nd 3rd STAR Reading[1]
Grades	K,1	1	2,3
Total N Baseline	231 (113 control)	277 (136 control)	633 (134 comparison)
Withdrawn T2 (Survey)	16 (6 control)	17 (4 control)	41 (10 comp, 9 late, 22 early)
Withdrawn T3 (Achievement)	24 (11 control)	28 (6 control)	85 (11 comp, 36 late, 38 early)
# Students Excluded Under Survey Criteria	0	1 (1 control)	3 (1 late, 2 early)
# Students Excluded Under Achievement Criteria	7 (4 control)	5 (1 control)	20 (17 comp, 3 late)
# Students Missing Survey Outcome Data	84 (14 int, 70 control)	90 (8 int, 82 cont)	128 (107 comp, 12 late, 9 early)

[1] Exclusion criteria for analysis: A change in between test 1 and test 3 outside of acceptable range. Exclusion criteria for just achievement analysis: Students needed to meet minimum hours, minimum weeks, and have a test 3 score with a test date.

Measures

Each student in the final analysis sample and many who were removed for the reasons mentioned in the previous section completed both a survey of student attitudes and an end-of-year assessment of reading or literacy ability.

Student Attitudes

The primary purpose of including an attitude assessment in the ARM impact study was to identify attitudes supporting higher academic achievement. An underlying assumption of the ARM impact evaluation is that students who have a positive attitude and feel happy when they read (Alexander & Filler, 1976; Mathewson, 1994; McKenna et al., 1995) will be more motivated to read. If students feel more motivated to read, they will achieve more success on academic measures.

A first step in identifying an assessment tool to measure the attitudes of children in Kindergarten to 3rd grade toward literacy was to explore the current literature. ARM's impact evaluation framework lead to the development of the following criteria for the ARM literacy assessment tool:

- *Focus on attitudes about literacy*
- *Administered with young children, Kindergarten-3rd grade*
- *Assessment tool with strong measures of validity and reliability*

Due to the fact that the academic literature on children's reading is vast, the evaluation team narrowed the scope of the available literature to studies about 1) children's motivation to read, 2) children's engagement in reading, and 3) children's attitudes about reading. For the impact evaluation, we focus our work on children's attitudes about reading and do not spend time exploring motivation or personal motivation to read. For that reason, only instruments developed to measure student "attitudes" were considered.

Additional criteria were added as the administration of the survey was identified, including a tool that could be administered by members with classrooms of children (rather than one to one) and a tool that could be scored by members. As a result, the ARM impact evaluation adopted the Elementary Reading Attitude Survey (ERAS). A copy of this survey is available at http://schools.liberty.k12.ga.us/jwalts/reading materials/Elementary Reading Attitude Survey.pdf.

Early Reading Attitudes Survey

The ERAS met our desire to focus on attitudes about literacy, had been administered with young children, Kindergarten-3rd grade, and was an assessment tool with strong measures of validity and reliability.

As a public-domain instrument, the ERAS is a frequently used instrument that may be used by teachers and school administrators to "(a) make possible initial conjecture about the attitudes of specific students, (b) provide a convenient group profile of a class (or a larger unit), or (c) serve as a means of monitoring the attitudinal impact of instructional programs" (McKenna & Kear, 1990, p. 628). The ERAS contains a total of 20 items, 10 focusing on recreational reading and 10 focused on academic reading. Students are asked to rate how happy or unhappy they are about a particular statement.

The ERAS has been used extensively by primary- and intermediate-level teachers to determine the overall attitude levels of classes. It has also provided insights into the reading habits and achievement levels of individual children. Unlike many affective scales, the ERAS exhibits solid validity and reliability characteristics, two critical attitudes given the potential importance of attitudinal indicators (Henk, Marinak, & Melnick, 2012). For example, in one study of more than 18,000 students, internal consistency coefficients for the recreational reading subscale ranged from 0.74 to 0.89. Adequate construct validity was reported by relations to library card possession and hours of reported television watching (McKenna & Kear, 1990).

The ERAS answer options are cartoon character Garfields with four different facial expressions. Each question is followed by 4 pictures of Garfield, a fictional cat in the Jim Davis comic strip, Garfield. While Garfield is age appropriate for ARM students, the impact evaluation advisory committee decided that the children were not familiar with Garfield and requested different answer options. The evaluators selected penciled kittens with four different facial expressions.

Academic Achievement

For Kindergarteners (and some 1st grade students), the achievement pre- and post-test was the STAR Early Literacy examination developed by Renaissance Learning. This assessment examined five domains in foundational reading skills and language: 1) print concepts, 2) phonological awareness, 3) phonics and word recognition, 4) fluency, and 5) vocabulary acquisition and use.

Students in the 1st, 2nd, and 3rd grades received the STAR Reading assessment. This assessment examined 46 reading skill areas in 11 domains and specifically targeted: 1) foundational skills, 2) literature (i.e., key ideas, craft and structure, and integration of knowledge and ideas), 3) reading informational text, and 4) language. For 1st grade students who were not able to exhibit an ability to read on the STAR reading assessment, assessment protocol quickly re-directed the student to the STAR Early Literacy examination, resulting in the directing of students into two analytic samples at this grade level. As a result, students who received an Early Literacy examination score on the pre-test were purposefully directed in the Early Literacy examination on the post-test.

Procedures

During the month of February, the evaluators trained all members on survey dissemination for students in grades K-3. We received permission for these members to

deliver the survey in one classroom within the building as well as to their assigned students during one-on-one or small-group tutoring time. During the months of February and March, all members proctored the survey using a tightly scripted administration protocol. They then entered student data on an entry-controlled spreadsheet and submitted the digital documentation to central administrators and on to the evaluators.

Following the survey, students received their end-of-year STAR assessment as a course of regular school day practice (as a reminder, all schools were required to use the assessment three times a year). The date of delivery of the final assessment ranged from February 11 to May 24, with nearly all submissions coming in April and May. Students tested before March 29 were not included in analysis due to a lack of an end-of-year assessment.

Other data collection

We collected other important data using three primary tools:

> *i. America Learns online portal*
> *ii. School Report*
> *iii. Administrative Records*

Members returned at least a monthly report of student tutoring hours and setting to the evaluators using an online platform called America Learns. Through this platform, they also provided reasons for not tutoring students when applicable and shared information with us on an end-of-year survey related to implementation adherence.

We relied on school report for all student demographic data, student-to-member assignment, student exit (such as moving), and student assessment results. Documents were exchanged with school site supervisors using a shared and password protected folder.

Finally, administrative records related to member quality, withdrawal, and reassignment were valuable to evaluators in record-keeping, decision-making regarding sample development, and as a moderating variable in secondary analyses.

Analytic Procedures

Student Attitudes

Students were first stratified by location, and post-hoc, grouped into three samples for analysis. The first group included all students who had not received intervention and any other students in the lower 1/3 of total intervention time at each building. The second group included all students who fit in the middle 1/3 of total intervention time at each building. The third group included all other students (i.e., the students who received the highest 1/3 of total intervention time). The preliminary stratification of students by building was necessary as a result of widely differing mean values towards reading at the building level. These rates are shown in Table 7.

Table 7. Distribution of ERAS subscale attitude scores across schools

School	Recreational Reading Attitudes	Academic Reading Attitudes	Mean
Amite County Elementary	30.5	31.4	31.0
Coffeeville Elementary	28.0	28.5	28.2
Eva Gordon Elementary	34.0	34.1	34.1
Fair Elementary	27.9	25.9	26.9
Franklin Lower Elementary	27.7	26.8	27.2
Goodman-Pickens Elementary	26.7	26.4	26.5
Hawkins Elementary	32.5	33.0	32.7
Holly Springs Primary	24.3	26.3	25.3
Jefferson County Elementary	19.5	19.0	19.2
Lumberton Elementary	27.4	27.8	27.6
Maddox Elementary	33.9	34.3	34.1
North Pontotoc Elementary	27.4	26.3	26.8
Oak Park Elementary	30.8	31.4	31.1
Okolona Elementary	30.2	29.9	30.1
Osyka Elementary	26.6	29.9	28.3
Petal Elementary	29.4	27.9	28.6
S.V. Marshall Elementary	27.4	28.8	28.1
Sherman Elementary	32.6	32.5	32.5
South Pontotoc Elementary	27.7	26.6	27.1
Warren Central Intermediate	24.9	25.1	25.0
Mean	**28.5**	**28.6**	**28.5**

The student attitudes analysis included students across all grade levels. In absence of a pre-intervention survey for student attitudes to show baseline similarity between groups, there was a need to use achievement results at t1. However, because not all students took the same assessment (some took STAR Early Literacy and some took STAR reading), there was a need to measure the percentile rank of each student. As the evaluators compared the mean percentile rank on the t1 assessment between the three groups, a small difference for one of the groups became apparent. Also, as shown in Table 8, the mean of total evaluation intervention time for each group varied widely as designed.

Table 8. Pre-test and intervention details for each student attitude evaluation sample

Analysis Group	N	Average Intervention Hours	Mean T1 Percentile
High intervention	243	57.0	27.0
Mid intervention	232	45.0	26.9
Low intervention	239	23.0	28.1
Mean		**41.7**	**27.3**

As a result, the student t1 percentile rank performance was included as a covariate in our student attitude analyses, along with intervention group. The evaluators completed the between-groups analysis using an analysis of co-variance (ANCOVA), a general linear model that blends ANOVA and regression.

Secondary Analysis. Following the primary analysis, a best fit model of additive covariates was included in a linear regression. These variables included those that described schools, members, and students themselves.

Academic Achievement

Our three samples of students developed for analysis of student reading achievement change were: a) Early Literacy, b) 1st grade STAR Reading, and d) 2nd-3rd grade STAR Reading. The study designs for two of these were experimental.

Experimental Study

Sample attrition met What Works Clearinghouse standards (U.S. Department of Education, 2013) for the Early Literacy group, as shown below in Table 9.

Table 9. Attrition rate of Early Literacy student achievement evaluation sample

Group	N	Attrition Rate
Control	98	13.3%
Intervention	102	13.6%

Sample attrition met with reservations What Works Clearinghouse standards for the 1st grade STAR group, requiring proof of baseline equivalence of the final analysis sample. Upon further review, the two analysis groups met requirements for baseline equivalence with a Cohen's D of 0.03. These details are presented below in table 10.

Table 10. Attrition rate and t1 mean achievement of 1st grade STAR student achievement evaluation sample

Group	N	Attrition Rate	Mean T1 Scale Score
Control	129	4.4%	69.2
Intervention	115	18.4%	69.1*

*Cohen's D = 0.03

Early literacy test takers (K and grade 1) were statistically similar with respect to these confounders. Preliminary analysis supports the use of an adjusted t-test to determine if the difference in improvement between students who received tutoring and control students was significant. Similarly, for 1st grade STAR test takers, an adjusted t-test was used in the final analysis, as this was supported by preliminary analyses. Additionally, residual values followed a random distribution, further supporting the use of a t-test (normality assumptions were met).

Quasi-Experimental Study

For the group of students in 2nd and 3rd grades, we used a priori propensity score matching to identify what became two intervention groups and a comparison group (see design section above). In an effort to meet, with reservations, What Works Clearinghouse standards, the evaluators sought to prove baseline equivalence.

Our three groups passed baseline equivalence tests, showing a maximum of 0.06 effect size difference between groups. As this value was not below 0.05, it was necessary for us to control for t1 in our analysis. These data are presented below in Table 11.

Table 11. Baseline equivalence of 2nd grade and 3rd grade STAR student achievement evaluation sample

Analysis Group	N	Mean t1 Scale Score
Early intervention	225	189
Late intervention	194	188
Comparison	106	193*

*Cohen's D = 0.06

Preliminary analyses support the use of a comparison test of the three groups based on a normal distribution (t-test) for all final analyses. Because the three groups for grades 2 and 3 did not meet base line equivalence for certain variables, these variables were included in the final analysis so that group means are adjusted for these potential confounders.

Secondary Analysis. Following the primary analysis, a best fit model of additive covariates was included in a linear regression. These variables included those that described schools, members, and students themselves. Models were developed for each of two samples, including students assessed with the: 1) STAR Early Literacy exam and 2) STAR Reading exam.

RESULTS

In this section, we present results of both primary and secondary analyses for each of our evaluation outcomes (student attitudes and achievement) and for each of our samples.

Student Attitudes towards Reading

We begin our presentation of results with our primary analysis of student attitudes. As mentioned in the "Methods" section, the research design resulted in the use of an ANCOVA for this outcome, investigating the differences in variance between three groups of students, while controlling for t1 percentile rank on the achievement examination. As a reminder, due to methodological limitations, students were not assigned to groups through randomization or matching. We present results for two analyses of student attitudes: 1) attitudes towards academic reading and 2) attitudes towards recreational reading.

Attitudes towards Academic Reading

As shown below in Table 12, our analysis compared student attitudes towards academic reading between three groups: those who received a high amount of intervention, those who received a medium amount of intervention, and those who received a low amount of intervention (or no intervention). Students within schools were spread across each of these groups to control for variability in mean school attitudes.

There was a non-significant effect among the groups ($F(3,710) = 2.20$, $p= 0.09$) when controlling for t1 achievement. The difference between the high-intervention and medium-intervention groups was positive towards the medium-intervention group and significant ($p= 0.02$). There was no significant difference between the attitudes of students in the high-intervention group and students in the low-intervention group when controlling for t1 achievement.

Table 12. Results from investigation of student attitudes towards academic reading

Sample	N	Mean t1 Percentile	Academic Attitudes Mean (SD)	Effect Size
high	243	26.95	2.70 (0.79)	---
mid	232	26.94	2.93 (0.69)	0.31*
low	239	28.11	2.86 (0.73)	0.21

* p<.05 ** p<.01 *** p<.001

Attitudes towards Recreational Reading

As shown below in Table 13, our analysis also compared student attitudes towards recreational reading between the three groups.

There was a significant effect among groups (F(3,710) = 2.91, p= 0.03) when controlling for t1 achievement. The analysis revealed a significant difference between the attitudes of students in the high-intervention group and students in the medium-intervention group towards the medium-intervention group in direction (p= 0.009) when controlling for t1 achievement. The analysis also revealed a significant difference between the attitudes of students in the high-intervention group and students in the low-intervention group in the direction of students in the low-intervention group (p= 0.02) when controlling for t1 achievement.

Table 13. Results from investigation of student attitudes towards recreational reading

Sample	N	Mean t1 Percentile	Recreational Attitudes Mean (SD)	Effect Size
high	243	26.95	2.73 (0.71)	---
mid	232	26.94	2.90 (0.68)	0.24**
low	239	28.11	2.91 (0.68)	0.26*

* p<.05 ** p<.01 *** p<.001

Student Achievement

We also present results from our analysis of student achievement change for our three groups of analysis (Early Literacy, 1st grade STAR, and 2nd and 3rd grade STAR). As explained in the "Methods" section, the Early Literacy and 1st grade STAR analysis concluded with a t-test between a randomly assigned intervention and control group. The 2nd and 3rd grade STAR analysis concluded with analyses of variance across three groups controlling for t1 scale score.

As shown in Table 14, the mean scale score change for the intervention group was lower than that of the control group in the Early Literacy evaluation group. Because we had built our analysis on a one-sided premise (as explained in the "Methods"), no further reporting was possible, as our expectation hinged on a more positive change score for the intervention group.

The results of the analysis of the 1st grade STAR evaluation group showed an insignificant but positive effect for the intervention group (t[242] = 0.53, p= 0.30).

The results of the analysis of the 2nd and 3rd grade STAR evaluation group were mixed. There was no significant effect for the intervention groups ($F_{[3, 521]} = 51.52$, $p > 0.05$) when controlling for t1 achievement. Again, because the mean change score for the early intervention group was lower than that of the comparison group, this made further analysis not applicable due to an assumption of a one-tailed change. There was a positive but non-significant effect of the late-intervention group as compared to the non-intervention (comparison) group ($p = 0.40$).

Table 14. Results from investigation of student achievement

	N	T1 Mean Scale Score (SD)	T3 T1 Scale Score Change (SD)	Effect Size
Early Literacy				
Intervention	102	482.4 (57.6)	224.0 (87.7)	N/A
Control	98	477.0 (55.0)	234.9 (97.3)	
1st Grade STAR				
Intervention	115	69.1 (3.1)	107.6 (83.5)	0.07
Control	129	69.2 (3.0)	102.1 (76.2)	
2nd-3rd Grade STAR				
Early Intervention	225	189.0 (74.5)	134.2 (69.2)	N/A
Late Intervention	194	188.0 (73.1)	138.9 (71.7)	0.05
Comparison	106	192.7 (75.2)	135.3 (78.0)	

* p<.05 ** p<.01 *** p<.001

Secondary analyses

Prior to completing secondary analyses for both the student attitudes outcomes and the student achievement change samples, we first loaded particular items on each of two scales (member quality, member setting report) onto factors using a principal component analysis. Please see Appendix C for the results of this analysis. Resulting factors were tested for internal consistency using Cronbach's alpha (Cronbach, 1951) along with results from our student attitudes survey. The resulting coefficients were: 1) Member Quality = 0.81, 2) Member Settings Report = 0.81, 3) Recreational Attitudes = 0.81, and 4) Academic Attitudes = 0.84. The coefficients for the two survey scales matched those reported by others (Kazelskis et al., 2005; McKenna, 2001; McKenna et al., 2012, 1995).

Second, we conducted a scatter-plot analysis of each potential independent variable that was a numeric integer against each of the dependent outcome variables. We saw no reason to infer any non-linearity in the relationships.

Third, we created a correlation table for all independent variables that were numeric integers, along with the outcome variables to determine collinearity among independent variables. These results allowed us to cluster certain variables as too highly correlated (r2 = 0.4 or higher) to include in the same model.

Fourth, we conducted a variance inflation factor (VIF) analysis on our resulting models. By doing so, we uncovered any collinearity between independent variables that were factors, or categorical in nature and all other independent variables (Lin & Lin, 2009). According to these authors, the VIF for any predictor is a measurement of its regression against the remaining predictors in a manner that results in values above 5 being evidence of collinearity. Our model returned just two variables as collinear: aggregated values for student academic and student recreational attitudes.

As a result, we investigated the reliability indices for the two attitude scales. The results showed a correlation of 0.93. This value was much higher than that seen in scale development literature, providing further support, in addition to VIF statistics, that the two scales were not independent for this sample. While the previous literature provided some justification for the existence of two separate factors, future work should leverage SEM with a large enough sample size to determine whether these two constructs are unique or if they reflect a more common general attitude toward reading.

Upon examining the two scale question sets more closely, the researchers decided to retain only school recreational attitudes for inclusion in the analysis. Because the dependent outcome in question is academic performance on English/Language Arts, it logically follows that student attitudes toward recreational reading, a construct that is heavily related to reading performance (Navarra, 2011), would be a more explanatory predictor than general attitudes toward academics.

Student Attitudes

A linear regression using information from multiple levels was calculated to predict changes in student reading attitude outcomes based on the differences in the magnitude of America Reads - Mississippi tutoring intervention received. The evaluation team developed two models with the change in academic reading attitudes as the dependent variable in model 1 and the change in recreational reading attitudes as the dependent variable in model 2. The pre-intervention academic performance as well as various demographic information (e.g., age, gender) describing the students were included to evaluate the true efficacy of the intervention. At the same time, we included member-level (member quality, member attendance) and school-level (member settings report, observed adherence rankings, reading proficiency rate, and aggregated reading

attitudes) co-variates to further control for effects at these levels. While we recognize the benefits of using multi-level modeling to determine the cross- and intra-level relationships among variables (Maas & Hox, 2005), there were too few sites and members to reliably analyze data using multi-level modeling (Aberson, 2011). Member data were fit by step-wise linear regression using both forward and backward directions (Venables & Ripley, 2002).

Results from the first regression model were statistically significant (academic attitudes) ($F[13, 511] = 6.98$, $p< 0.001$), with the model accounting for approximately 12.9% of the variance in academic attitudes. Results from the second regression model were also statistically significant (recreational attitudes) ($F[13, 511] = 7.69$n $p< 0.001$), with the model accounting for approximately 14.2% of the variance in recreational attitudes.

Table 15 presents the outcomes of the analysis specific to all students on each of two measures (academic attitudes and recreational attitudes). The number of intervention weeks was positive and significant for academic attitudes ($p< 0.001$). Conversely, the number of intervention minutes was negative and significant, though with little effect, for both academic attitudes ($p< 0.001$) and recreational attitudes ($p< 0.001$). A number of student-level variables did prove significant, including male gender for both academic attitudes ($p< 0.001$), and recreational attitudes ($p= 0.004$) and reported issues during tutoring for recreational attitudes ($p< 0.001$).

At the same time, a number of member-level variables were significant, including positive effects for member attendance for both academic attitudes ($p< 0.001$) and recreational attitudes ($p= 0.01$) and member discipline status for both academic attitudes ($p= 0.03$) and recreational attitudes ($p< 0.001$). An additional member-level variable, member quality, was significant and negative for academic attitudes only ($p= 0.004$).

In addition, school-level variables were significant, including positive effects for observed adherence rating for both academic attitudes ($p< 0.001$) and recreational attitudes ($p= 0.001$). An additional school-level variable, the member report on their tutoring setting, was significant and negative for both academic attitudes ($p= 0.005$) and recreational attitudes ($p= 0.05$).

Table 15. Results from Regression Model Predicting Student Attitudes (n=525)

Variable	Outcome	
	Academic Attitudes	Recreational Attitudes
intercept	0.51	1.02
Intervention Weeks	0.05***	0.02
Intervention Minutes	-0.00***	-0.00***
Student-Level Factors		
Male	-0.21***	-0.17**
Older than peers[1]	0.13	N/A
Younger than peers[1]	0.50	N/A
African-American	-0.12	-0.28
White	-0.02	-0.26
Pre-Test Achievement	0.00	0.00
Reported Issues Tutoring	N/A	0.13***
Member-Level Factors		
Member Quality	-0.28**	N/A
Member Attendance	0.57***	0.37*
Member Discipline Status	0.37*	0.55***
School-Level Factors		
Member Settings Report	-0.20**	-0.14*
Observed Adherence Rating	0.05***	0.04**
State Reading Rating	N/A	0.07

* p<.05 ** p<.01 *** p<.001

[1]*Determined from age range variable*

Student Achievement

We conducted two linear regressions using information from multiple levels to evaluate effects of the differences in the magnitude of America Reads - Mississippi tutoring intervention received. In these models, the change in scale score on an achievement test was the dependent variable (either the Early Literacy examination or the STAR Reading examination; see below). The pre-intervention academic performance as well as various demographic information (age, gender) describing the students were included to evaluate the estimated efficacy of the intervention. At the same time, we included member-level (member quality, member attendance) and school-level (member settings report, observed adherence rankings, reading proficiency rate, and aggregated reading attitudes) co-variates to further control for effects at these levels.

Member data were fit by step-wise linear regression using both forward and backward directions (Venables & Ripley, 2002).

Early Literacy

Results from the regression predicting changes in scale scores on the early literacy assessment were statistically significant, ($F_{[10, 62]} = 4.67$, $p < 0.001$), with the model accounting for approximately 33.8% of the variability in the outcome.

Table 16 presents the outcomes of the analysis specific to students tested using the Early Literacy examination (grades = K, 1). The intervention minutes were positive to a point, and then negative as the number of minutes increased towards a cluster of high intervention. Neither finding was statistically significant. A number of student-level variables did prove significant. Member-level and school-level variables were not significant.

Among student-level variables, only those related to student age were significant. Showing as positive and significant were status as outside of and older than the 12-month range standard for a grade level (i.e., the student was held back a year) ($p = 0.01$), as well as relative age within grade ($p < 0.001$). The results also show a negative and significant relationship to months of age in general ($p < 0.001$).

Table 16. Results from Regression Model Predicting Early Literacy (n=73)

Variable	Early Literacy Change In Scale Score
Intercept	695.5**
High Intervention Minutes	-38.3
Moderate Intervention Minutes	16.4
Student-Level Factors	
Months of age	-9.35***
Older than peers[1]	102.0*
Younger than peers[1]	18.2
Age offset (relative to peers)	13.1***
Pre-Test Achievement	-0.48
Member-Level Factors	
Member Quality	-46.5
School-Level Factors	
Member Settings Report	39.3
Observed Adherence Rating	5.7

* p<.05 ** p<.01 *** p<.001

[1]*Determined from age range variable*

STAR Reading

Results from the regression predicting change in scale score on the STAR Reading assessment were statistically significant, (F[15, 442] = 5.49, p< 0.001), with the model accounting for approximately 12.9% of the variability in the outcome.

Table 17 presents the outcomes of the analysis specific to students tested using the STAR Reading examination (grades = 1, 2, 3). The number of intervention weeks was slightly negative but not significant. A number of student-level factors did prove significant, including starting pre-test achievement, which findings showed had a negative relationship (i.e., those with the lowest starting scale score improved their score more) (p=0.001). Status as outside of and older than standard-aged peers was also negative and significant (p= 0.03).

In this case, member-level attendance and school-level factors (reading achievement rate, aggregated reading attitudes) also proved positive and significant. Thus change in

scale score on the reading achievement increased as a function of positive changes in these covariates (i.e., better member attendance, more positive reading achievement rates, or student attitudes towards reading).

Specifically member attendance (p< 0.001), reading achievement rate (p= 0.03), and aggregate reading attitudes (p= 0.03) were significant.

Table 17. Results from Regression Model Predicting STAR Reading (n=458)

Variable	STAR Reading Change In Scale Score
Intercept	-378.6***
Intervention Weeks	-1.3
Student-Level Factors	
Male	12.1
Older than peers[1]	-24.0*
Younger than peers[1]	36.4
Age offset (relative to peers)	-1.1
Pre-Test Achievement	a) -34.7***, b)-23.4*, c)-33.8**, d) -41.4***
Member-Level Factors	
Member Attendance	77.4***
School-Level Factors	
Member Settings Report	14.5
Observed Adherence Rating	2.0
3rd Grade Proficiency Rate	0.9*
Aggregated Reading Attitudes	2.2*

* p<.05 ** p<.01 *** p<.001

Pre-Test Achievement categories: a) Highest Achievement Group, b) High Achievement Group, c) Moderate Achievement Group, d) Low Achievement Group; Lowest Achievement Group set as intercept

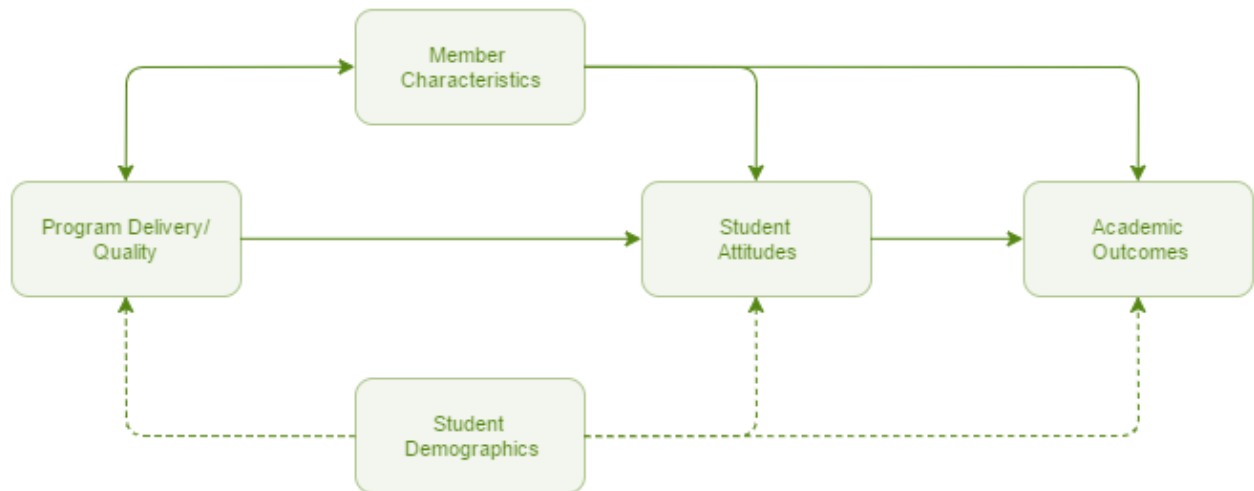[1]*Determined from age range variable*

Commonality Coefficient Interpretation

We sought to better understand the relationship among the variables included in the study and the potential impact they may have had on student academic or attitudinal outcomes. Thus, we conducted a commonality analysis (Nimon, Lewis, Kane, & Haynes, 2008; Nimon & Oswald, 2013), which attempts to explain the independent and shared predictive power of predictors in regression models. Our previous steps included the calculation of regression coefficients that seek to determine the strength and direction of the relationship of many variables in relation to the outcome of interest (e.g., STAR achievement, early literacy achievement, student attitudes toward academics, and student attitudes toward recreational reading). However, it is likely inaccurate to assume that the variables examined are completely independent of one another—a key assumption in regression analyses.

Background and Interpretation

The following analysis allows us to better understand a predictor's contribution to an outcome of interest in isolation and in combination with other predictors (Nimon, 2010). In other words, this analysis allows us to identify the areas that program staff should focus on to yield the most significant improvements in student outcomes given a limited set of resources. The following findings are presented to attempt to conceptualize the relative value of future efforts and how that relates to previous findings.

Before examining the way in which the predictors contributed to the outcomes, it is important to first name the possible theory of action behind the regression and CA. It is expected that there are many characteristics of the program, students, and environment that contribute to student performance on either the STAR Reading or Early Literacy assessment**. However, it is important to identify which characteristics or variables might be intermediate steps on the way to the student outcomes. A high-level theory of action is presented below:**

The proposed high-level theory of action is intended to help contextualize the CA coefficients and percentage of variance explained by the unique and combined predictors. As a point of clarification, not every predictor was included in the CA because every predictor increases the complexity of the model and its results exponentially. For example, if 2 variables are used, then a possible 3 results exist (V1, V2, and V1*V2). If 3 variables are used, then a possible 6 results exist (V1, V2, V3, V1*V2, V1*V3, V2*V3). Continuing this, 4 variables result in 24 possible combinations, 5 result in 120, and so on. Therefore, we strategically selected those variables that were the strongest predictors in the previous models using the following steps: 1) We selected the independent variables that were significant predictors at the .05 level or better, then 2) Filled out to five variables for CA if room existed.

Results

For detailed results on each of the four individual analyses (Academic Attitudes, Recreational Attitudes, Early Literacy, STAR Reading), please see **Appendix D: Commonality Coefficient Analysis**.

Here, we will present an overview of findings across all four analyses.

As noted in the previous section, the linear model for each outcome differs noticeably, which presents some difficulty in looking across CA attempts for more generalizable findings. However, we did indicate a number of variables that either through their own unique predictive power or through a common predictive power shared with another variable in the same model, did show similar values across tests. Table 18 below shows all such variables with at least ten percent predictive power for one of the models.

Table 18: Commonality Analysis coefficients for variables with 10% or more predictive power for at least one outcome

Variable	Academic Attitudes	Recreational Attitudes	Early Literature	STAR Reading
Unique to Age in Months			72%	
Unique to Age Offset			35%	
Unique to Age Range			16%	14%
Unique to Gender	12%	9%		
Unique to Member Attendance	10%			33%
Unique to Member Quality	3%		19%	
Unique to Number of Weeks	10%			
Unique to Observed Adherence	26%	27%	5%	
Unique to Pre-Test Achievement Group				30%
Unique to Reported Issues Tutoring		20%		
Unique to Total Intervention Minutes	57%	46%		
Common to Age in Months, Member Quality Factor			-10%	
Common to Age Offset, Age in Months			-33%	
Common to Member Attendance, Total Intervention Minutes	12%			
Common to Observed Adherence, Age in Months			14%	
Common to Reported Issues Tutoring, Total Intervention Minutes		19%		
Common to Total Evaluation Minutes, Observed Adherence	-18%	-17%		

At first glance, a few variables stand out. First, are a group of variables that predict a good deal of what we know in one of our models, for example, the variable age in months as a 72 percent predictor towards Early Literacy achievement. Second are variables that predict some of what we know about more than one of our models. An example of this type of variable is observed adherence, which predicts towards models for academic attitudes, recreational attitudes, and Early Literacy outcomes. Also worth noting are the positive values for total intervention minutes as related to student attitudes when commonality interferences are stripped away and only the unique predictive power is exposed. In the context of the findings in the model, it appears it is this interference that drives the net effect from being ever so slightly positive to nearly zero but negative.

DISCUSSSION

Through a randomized controlled trial design and careful data collection and exchange, we were able to investigate the America Reads - Mississippi program's impact on student academic gains in reading during the 2015-2016 school year and the impact of the same program on student attitudes towards academic and recreational reading. Unfortunately, our primary analyses did not provide evidence that the program provided an additional impact to the impact of general classroom instruction on either student achievement or student attitudes towards reading. Through secondary analysis, we were able to isolate for a number of student-level, member-level, and school-level variables that predicted positive changes in outcomes. However, even when controlling for these variables, the intervention-related variables of dosage still did not prove to significantly impact positive change in either achievement or attitude outcomes.

In order to fully understand our findings, it is important to place them in the context of previous studies reviewed in the introduction. For example, the lack of a significant effect of the tutoring intervention in the current study actually corroborates the findings of other studies that only found significant positive effects for foundational skills but not higher-order skills and that reported that implementation adherence was central to supporting potential outcomes.

Failure to Generate Positive Significant Effects

A glance at the existing literature informs us that no matter the success in implementation of the program in practice, it is likely that the tutoring intervention provided by America Reads - Mississippi members would not have met the needs of every student. For example, O'Conner (1997) explained that 10 percent of the students receiving a one-on-one intervention in 1st grade failed to benefit from their selection for the intervention group. Al Otaiba and colleagues (2005) also mentioned that many children in their study did not benefit from tutoring. These findings were corroborated by Vandervelden and Siegel (1997), who reported that 20 percent of their lowest-scoring Kindergarteners showed no improvement.

Vellutino et al. (1996) found that 33 percent of the intervention sample still scored below the 30th percentile on standardized tests at the end of the intervention. Other authors (Gómez-Bellengé, 2004; Marulis & Neuman, 2013) have also noted the inability for research-based tutoring programs to contribute a similar effect size for students of low socio-economic status as that for other students.

Even Vadasy and colleagues (2000) noted that five students (22 percent) of the study sample still scored in the lowest percentiles in reading following that study's tutoring intervention. The author's argument was that even with successful implementation of effective tutoring programs and practices, some children will still require more intensive and longer assistance than this program provided or different assistance altogether.

In fact, one author suggested that lack of a reaction to early tutoring should actually serve to identify children who require more expert and intensive instruction (Vellutino et al., 1996). For example, Vadasy et al., (2000) stated that Title 1 teachers and special education teachers supporting his study regarded a student's failure to improve in a tutoring program as an indicator for potential special education assessment. This argument actually aligns with the structure present in the Minnesota Reading Corps (Markovitz et al., 2014), as the members in that program serve students in Tier 2 settings of the Response to Intervention (RtI) framework in each school. In this tier, members provide assistance to students with demonstrated need, and those who do not make satisfactory gains would move on to Tier 3 interventions, including individualized attention and potentially the development of an Individualized Education Plan (i.e., disability status).

In the case of the current study, most tutoring was performed in grade levels (K-2) where the identification and documentation of disabilities occurs at a lower rate than in tested grades. Though the program is not structured to identify students with disabilities, the reaction of students to member service in pull-out small-group tutoring can certainly provide documentation of student need for more individualized and professional services.

One consideration that has been suggested is to extend offering interventions across school years. A number of scholars have suggested that interventions of more than one year may be needed by some children, such as those of low socio-economic status (Juel, 1996).

Less Than Desirable Fidelity of Implementation

This particular study design did not isolate for effects of training and supervision on members and local sites. However, previous research (Vadasy et al., 2000) found that providing more training in lesson components before tutors began working with children, along with increased supervision, resulted in more accurate implementation. In fact, the author argued that although tutors can provide critical instruction to struggling readers, they must first receive considerable training, support, and supervision. As just such an example, he referenced his own earlier study that had less intensive supports and less positive outcomes (Vadasy et al., 1997).

The findings of Elbaum and colleagues (2000) substantiated these results, as in their meta-analysis, the authors reported an average effect size of 0.59 for studies that mentioned the

training of volunteers, compared to just 0.26 overall for studies of volunteers. In the case of the America Reads - Mississippi program, member training was minimal compared to many studies referenced in a review of literature. A lack of resources limited the program training to just two in-person training events and nominal supervisory meetings. However, even at this minimal level, secondary analyses uncovered a positive and significant effect related to member attendance at these required trainings and supervisory meetings.

To sustain continued acceleration of tutor skills outside of program-provided training, Al Otaiba and colleagues (2005) stated that it was "critical for programs" to have tutors plan interventions and resolve issues in lockstep with school leadership and classroom teachers. Unfortunately, in our study, local site ratings during multiple observations of program adherence showed highly variable site supervisory effectiveness, a finding that produced significant positive impacts during secondary analysis (i.e., sites with tighter implementation adherence and more structured supports saw more positive changes). For example, many America Reads - Mississippi members reported that they did not have access to even monthly support meetings with a paid professional at their site or had no curricular support from specialists or classroom teachers.

In general, other studies have noted the difficulty that some individual schools face in bringing researched practices to scale. For example, one report (Vadasy et al., 2000) noted that the building must include in its supporting infrastructure "individuals who can recruit reliable and conscientious tutors, provide training and supervision, integrate tutoring into the schools' schedules, give technical assistance on instructional and management problems, and help in assessing student progress." From our experience, many schools could not devote the resources required to complete these tasks. In addition, a number of schools did not provide tutoring from students in the months of September and October, and even into November, because of difficulty scheduling sessions and pairing members and students, a finding shared by Al Otaiba and colleagues (2005).

To this end, the current study seemed to confirm the necessity of pre-intervention training for all non-professionals who desire to tutor students in a school setting, as shifting training responsibilities to buildings was unsuccessful. One team of scholars warned that this desire for more training up-front must be balanced with the fact that a low training hurdle will serve as an attraction to potentially interested adults for an unpaid position (S. Baker, Gersten, & Keating, 2000). However, in the case of America Reads - Mississippi, members were required to volunteer for an entire year's worth of volunteer service, which is a much more pragmatic hurdle itself.

No matter the extent of program-provided training, it is imperative that local sites continue to advance the skills of their embedded volunteers by including them in all professional learning opportunities that would be offered to paid educators at the school while also providing an engaged site supervisor to provide even more skill and curriculum development. At the same time, current results reinforce the need to consistently monitor implementation within local buildings so program staff can intervene when necessary in poor-scoring sites.

The America Reads - Mississippi program also faced a difficult and unexpected challenge during the school year studied: most of the program's members were new, and most, if not all, of the members who did not return had successfully completed their maximum two years of service. There were also, however, a small number of early exits due to career opportunities, exits for personal compelling circumstances such as a medical issue or relocation, or early exits due to justified administrator concerns. Member retention is certainly important to consider as related to tutor development. Many studies have referenced the positive impact of the continued development of tutors across multiple years (Invernizzi, Rosemary, Juel, & Richards, 1997; Vadasy et al., 2000).

Our findings serve to extend the findings of the current body of tutoring program research related to a number of program and evaluation features, including 1) the ability for volunteer tutors to successfully navigate effective intervention sessions without a standardized curriculum, 2) the influence of the tutoring intervention on student attitudes, 3) the ability for volunteer tutors to produce positive and significant effects for Kindergarten students, 4) the measurement of comprehension as an intervention outcome, and 5) the efficacy of volunteers supporting student comprehension efforts.

A Potential Need for a Standardized Curriculum

The America Reads - Mississippi program does not rely on a standardized curriculum but instead relies on the close interaction between teachers, site supervisors, and tutors in curricular decisions. Vadasy and colleagues (2000) suggested that though such a cooperative program (in coordination with classroom reading instruction) would be desirable, the practical application of such a setting is virtually impossible due to tremendous diversity across classrooms in what is being taught and how the content is being delivered. Another study (Invernizzi et al., 1997) argued that unless the classroom teacher can personally adjust the tutoring lessons of each child's lesson to that child's classroom progress, a standardized approach is likely the best approach when other individuals are responsible for tutoring.

The inability of many members to find lockstep with site supervisors and classroom teachers certainly reinforced prior findings in this particular setting. In many cases throughout the school

year studied, members informed us that they were receiving no curricular materials at all from school staff in the building or had been given a large binder of resources at the beginning of the year without supervision on how and when to use particular lessons. In fact, some members even reported accessing the Internet on their own to try and construct lessons to use with students. Even so, one study (S. Baker et al., 2000) of a similarly structured program reported significant positive effects on students in lieu of a standardized curriculum (though even this study failed to show significant results for passage comprehension).

No Demonstrated Positive Effect on Student Reading Attitudes

As mentioned in our review of literature, there are scarce examples of authors investigating the impact of tutoring on student attitudes towards reading. This is perhaps a result of the difficulty in delivering pre-test and post-test measures to both the intervention and non-intervention groups (as we describe at length later in the limitations section). Maybe this area will be one of the next that scholars investigate. It is certainly an area of interest, as a number of authors have documented a link between aspects of attitude, such as motivation, to greater achievement or other factors affecting achievement (Elley, 1992; Gambrell et al., 1996; McKenna, 2001; McKenna et al., 2012).

Given our ability to deliver a post-test primarily to those students who received the intervention, we can only infer from the results of this study that those students who had received the most tutoring felt less positive about both academic reading and recreational reading at the time of testing (February and March). As mentioned above, the secondary analysis controlled for numerous other variables and explored unique predictive power to each variable and shared effects between variables. Through these analyses, we determined that the effect of the amount of tutoring was significant but negligible (with the direction of that effect being mixed depending upon the analysis).

The only previous study that examined the impact of a similar program on student attitudes was that of Pena (2008) about a lunch-time tutoring program in Chicago schools. Both Pena's study and the current study (albeit in a secondary analysis) found a significant positive effect on just one of the attitude scales. In the case of his study, that scale was the recreational attitude scale (and was driven primarily by a positive increase for girls—a gender difference noted in our study as well). However, the author stated that the program in this study targeted recreational reading specifically. In addition, Pena's design allowed for the evaluation of sustaining impact across multiple cohorts, for which there was no significant effect for either academic attitudes or recreational attitudes. In fact, these results substantiated the evidence around the deterioration of student attitudes towards reading as they age (Lazarus & Callahan, 2000;

McKenna et al., 2012, 1995), even within one school year. We were unable to comment on this phenomenon due to our post-test-only design.

Pena (2008) also studied an additional hypothesis, finding that differences in attitudes could not predict student achievement, whereas our study reversed the independent and dependent variables and found that pre-test achievement could not predict later attitudes. This finding is of interest in relation to results from our study, which showed mostly non-significant but sometimes positive impacts of the intervention on achievement with mixed and significant impacts on student attitudes (even if very small). Secondary findings also showed that implementation adherence observation scores were a positive and significant predictor of reading attitudes (i.e., the tighter the tutoring to the program design, the most positive attitudes of the students).

These findings seem to indicate that the relationship between the intervention, individual student attitudes, and student achievement is overwhelmingly complex, an inference supported by other findings (L. Baker & Wigfield, 1999; Wigfield & Guthrie, 1997). To further complicate matters, in the current study's overall aggregated findings on attitudes towards recreational reading was also a positive and significant predictor of reading achievement among 1st grade through 3rd grade students. Thus, while like others, we did not find an association between student attitudes and motivation achievement (L. Baker & Wigfield, 1999; Pena, 2008; Wigfield & Guthrie, 1997), we did gather preliminary evidence that general school-wide positive attitudes towards reading might have a small but significant effect on student achievement. However, we found no other studies with which to authenticate this finding.

Volunteer Effects on Kindergarten Students

The America Reads - Mississippi tutoring program as evaluated in the current study was unable to impart any positive gains to non-reading younger students in Kindergarten and 1st grade. In reflection on this finding, we were reminded of the results of Al Otaiba and colleagues (2005), which showed that Kindergarten students needed a significant dosage (thirty minutes, four days a week, for an entire school year) to achieve a positive and significant effect. In the same study, the authors reported no significant effect for Kindergarten students provided just two days of tutoring a week for the entire school year.

In the current study, the intervention was typically provided either less than four days a week or for about half a school year (or both). The current study may serve as additional evidence that tutoring programs must be extensive in dosage for this population of students, starting immediately with the beginning of the school year and lasting through the end of the school year (or as previously mentioned through to a second year) (Juel, 1996).

In this case, program administrators had only recently restored Kindergarten as a selection option for member service at the request of the schools it serves. A few years earlier, following a previous evaluation, administrators had sponsored only member tutoring for students in 1st grade and above, as they expected that members would not have the skills to support students who could not yet read. Given the results of the current study and its corroboration of previous evaluation findings, one could argue that perhaps members should focus their activities on struggling readers rather than non-readers.

The Measurement of Comprehension

The only achievement measurement instrument available to the program was a set of two tests of more generalized reading ability that included a major component of comprehension. However, previous studies have often reported non-significant findings related to higher-order reading skills, even when reporting significant positive effects on foundational skills. In one study (Vadasy & Sanders, 2008) the effect size for a study of reading comprehension was just 0.09. Also, while the findings of Wanzek and colleagues (2015) indicated moderate, positive effects on foundational reading skills, they found much smaller reported effects on measures of language/comprehension. Further, most studies included in their meta-analysis relied on professional tutors rather than volunteers.

A study by Osborn and colleagues (2007) presented a tutoring intervention that had averaged more than a month's worth of gains for each month of intervention on some foundational skills. However, significant effects were not attained for a broad reading measure that included the Woodcock Johnson III Letter-Word Identification, Reading Fluency, Passage Comprehension, and Word Attack instruments, measuring comparable constructs to those assessed on the achievement instruments used in this study. Even more, the students in this study received their intervention from paid professionals (the study references classroom teachers, paraprofessionals, and administrators) rather than volunteer tutors.

The current study attempted to measure that impact on exactly the types of measures for which previous studies had found only small or no effects, and attempted to achieve these impacts with volunteer rather than professional interventionists. To this end, the current study contributes to the conversation around volunteer tutors and projected impacts, as other studies promoting significant positive effects of similar programs typically investigated outcomes on the foundation skill level rather than at the higher-order level of language/comprehension.

The Efficacy of Volunteers Supporting Comprehension Efforts

As mentioned previously, the America Reads - Mississippi program itself does not provide extensive training on the pedagogy of supporting students' foundational reading skills, but rather asks the local sites to assume that responsibility. In practice, we often observed members leading small group reading circles or asking questions about comprehension following a silent reading passage rather than supporting foundational skills.

In a way, because the study results reflect evidence that local sites did not consistently provide the desired level of pedagogical support for foundational skill development and because our STAR instrument was a measure of reading comprehension, we see the findings as an extension of the findings of Torgeson, and colleagues (1999), who found that focusing on skills like reading fluency at the expense of phonological skills will not necessarily result in positive changes in reading comprehension. This argument is echoed by a number of authors (Kershaw & Schatschneider, 2012; Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009) who wrote that a focus on phonological reading skills first will then facilitate reading fluency and accuracy, and in turn, reading comprehension. As a result, it seems imperative that all members work first with students on these foundational skills before progressing to fluency and comprehension-type tasks.

Even then, Vadasy and colleagues (2000) argued that if volunteers can successfully help at-risk students develop foundational phonological reading skills, then the expertise of more highly trained professionals, rather than more volunteer time, might be needed to extend this skill advantage towards comprehension. Considering the theoretical implications, it is not surprising that most tutoring programs with a standardized curriculum often focus only on foundational skills (Allor & McCathren, 2004; Y. S. Lee et al., 2010; Osborn et al., 2007; Vadasy et al., 1997; Wanzek & Vaughn, 2007). In fact, each of the tutoring programs Wanzek and colleagues (2015) included in their recent meta-analysis studied interventions that focused on foundational reading skills or multi-component activities.

The Determination of the Correct Dosage

There has been considerable variation in the literature as to the degree to which studied programs were dosed when delivered to students. Though theoretically it might make sense that more of the intervention is better for all students, one also needs to consider the opportunity cost of preventing the student from accessing other learning experiences (perhaps with certified professionals) when pulled out by a volunteer for tutoring.

In fact, two teams of scholars completing meta-analytic surveys of the recent literature established no effects of additional dosage on the reading outcomes for students in studies investigating effects on mostly 1st - 3rd grade children (Tran, Sanchez, Arellano, & Swanson, 2011; Wanzek et al., 2015). These authors suggested that this finding was evidence that foundational skills may be positively affected in a short amount of time. We would contend that the findings of the current study, specifically those related to the lack of a significant effect of intervention dosage on student achievement, be considered in the context of the findings reported by these authors with one caveat. In this case, we might infer that the findings of these two studies authenticated by the current study suggest that it might be difficult if not impossible to overcome any breakdown in implementation fidelity through the provision of more weeks, days, or hours of a tutoring intervention. In other words, having more of an intervention that is not working does not lead to more positive outcomes.

However, our findings do conflict with many published previous findings. Among published reports, there was ample evidence of volunteer-led tutoring efforts producing significant positive effects for tutored students. For example, the findings of our study contrasted with several impact studies of volunteer tutoring with similar treatment intensities, settings, and scope. These include 1) the Vadasy and colleagues (2000) study of first grade students receiving 30 minutes of tutoring at least 4 days per week, which showed effect sizes of at least 0.42, 2) the significant and positive findings of Lee and colleagues (2010) related to the Experience Corps program, and 3) the evaluation of successful implementation of the two-year SMART program using adult volunteers and minimal training to tutor 1st - 2nd grade students (S. Baker et al., 2000). However, it is important to consider that most authors do not report on effects that are not positive or are not significant, and thus findings similar to those of the present study are unlikely to be found in the literature (Rodrigues, 2013).

Limitations

Returning to the findings of Wanzek and colleagues (2015), the authors determined that variance among studies was not significantly explained by intervention type, instructional group size (if five or less students), grade level, implementer, or total hours of intervention. As a result, one might argue that the research supports an interpretation that a severe limitation in this study was the lack of implementation adherence at the local level. An additional associated issue was the lack of validated tools and travel funding to allow for the collection of data either in person or reported by individual members.

During our somewhat limited observance (between one and four times during the year), program administrators certainly recorded enough evidence to question whether any lack of impact should be attributed to the tutoring program or to variable program adherence at the

school level. For example, the program did set structures and guidelines according to known theory along setting lines (e.g., group sizes, grade levels, hours, and days). However, decisions in some individual buildings prohibited the members from implementing their tutoring efforts in strict accordance with their America Reads - Mississippi mission. Administrators were certainly attentive to these challenges when noted and made ongoing efforts to address known issues in an effort to ensure effectiveness and adherence to mission. However, time that had already lapsed could not be recovered through an extension of tutoring time per week, or an extension in the number of weeks of programming.

It was clear that not all buildings could afford the space to implement the program as designed. However, when faced with resource challenges, not all locations acted creatively to develop a space that could at least support positive outcomes. For example, many members were only given space within the classroom to deliver their pull-out intervention. During our visits, we saw that these environments could be supportive settings when they were moved off to the side and separated from the classroom by some kind of structure, like a bookshelf, but we also saw tables that sat in the middle of the classroom with no discernible barrier creating a quiet and private space for tutoring.

We also witnessed tutoring at one school happening in the school gymnasium, a large echoing space that was also being used by other children for play. Though administrators assured us that this space was not typically used for tutoring, we suspect that at other times in that building and at other locations, a space crunch forced tutoring into wildly un-supportive environments. This limitation was mentioned by Al Otaiba and colleagues (2005), who witnessed tutoring occurring after school in a school lunchroom while other students were practicing for a school performance.

Finally, on more than one occasion, we found "small groups" that included eight or more students crowded around a table. Perhaps the lack of fidelity around local "small group" definitions was one of the reasons that other authors advocated for the use of only one-on-one settings when volunteer tutors deliver interventions (Askew & Simpson, n.d.; Chang, 2011). It should be noted that Chang (2011) made this argument following a study of an intervention delivered by college students who had already received a fair amount of teacher training. One can only expect that the author might further prioritize a one-on-one setting when the typical tutor had received no teacher training and more than likely will have never been an active educator in any school classroom as with the program in the current study.

The removal of school and member choice about one-on-one versus small groups would certainly help to ease fidelity concerns related to both effective group sizes (i.e., schools choosing a more is better approach and loading up tables) and pedagogical concerns (both in

the behavior control of large groups of students, but also in the preparation of practice materials for students with different degrees of need). In addition, the decision would certainly parallel much of the research highlighted in our review as well as a previous meta-analysis (Elbaum et al., 2000) and suggested practice by the federal government (Institute of Education Sciences, 2003, p. iii).

At the same time, future research efforts should focus on the development of tools to more effectively capture the adherence to planned program structures. These instruments could assist in direct observation as well as the data collection from professionals in affected buildings from the members and even from the students (as we did with the survey of attitudes in this study). As mentioned previously, our principal component analysis of existing tools showed that individual items on surveys and rubrics were not loading onto expected components. The continued development of these tools as outlined in the appendices will certainly assist administrator, evaluator, and other stakeholder inferences related to implementation. Such tools will partially alleviate the additional challenge faced by program administrators around the time-consuming and costly efforts dedicated to collecting in person the data necessary to ensure that tutoring environments are conducive to student learning.

Ethics Around Student Selection

We do think it is important to share that our evaluation found no differentially effective impact of the program on any subgroups of students, such as those of ethnic minority status, ELL, or low socio-economic status. The only significant effects identified in secondary analyses were for gender (male) and age (especially relative age to same-grade peers). These results provide evidence that the program should not target demographic-type information that educators know about students, confirming the findings of other authors (Y. S. Lee et al., 2010). Rather, tutoring program administrators informed by this study should look to target students based solely on need.

Methodological Limitations

There were also methodological limitations that affected our ability to measure impact, but also reinforce the need to avoid generalizing these results. First, just a small sample of students in Kindergarten participated in the study due to school choice to focus member attention on other grade levels. The small sample size caveat was also present in 1st grade due to use of two different assessments for students in this grade level. Due to ethical considerations, evaluators and programmers recommended that school personnel make the decision as to whether the student would take the Early Literacy or STAR Reading assessment. Unfortunately, because

these two assessments test starkly different constructs, these students needed to be separated during analysis.

Unfortunately, the agreements supporting the partnership between the local school buildings and central program administrators did not identify the assessment of student attitudes as part of the contract. As a result, the evaluation team and central administrators were advised by a committee of site administrators to disseminate any attitude measures only once during the year, and primarily only to those receiving the intervention. As a result, the evaluators were unable to test differences in student attitudes between groups of intervention and non-intervention students. Even if it had been possible, we would have been unable to establish baseline equivalence between our samples due to the lack of a pre-test.

Instead, we created three groups based on the relative exposure to the intervention within each local building and used pre-test student achievement as a proxy for attitudes in an effort to establish equivalence for our non-experimental analysis. These issues deeply impact any inferences that should be drawn and likely made it more difficult to establish a real effect where we saw significant results. In future evaluations of project impacts, we will ensure the delivery of student attitudes measures to students both before and after the completion of tutoring, even if only for intervention students.

As mentioned earlier, many of the variables used during our secondary analysis were the result of data collected using tools that were not built using a methodical and scientific approach, but rather were developed quickly in order to collect data where no existing tools or validated tools existed. At best, these tools should have been considered for pilot in a number of schools before statewide use. However, given the nature of the evaluation and mapping onto school schedules, this was not possible. Future research efforts will better measure program areas, especially measuring implementation at the local level.

It must also be mentioned that the theory behind the America Reads - Mississippi program includes outcomes at multiple levels. Rather than just a focus on effects on student attitudes and achievement, the program is also tightly aligned to both member-level and community-level outcomes. For example, the program sees itself as a training ground with the capacity to entice potential educators to finish the required training in order to become a licensed professional. At the same time, it may provide the experience necessary for those who might struggle in the profession to consider alternative careers before being tasked with the leadership of their own classroom. In fact, this member-level effect was also noted by Chang (2011), who saw the program evaluated in that review as an "opportunity for teacher candidates to explore their passion for teaching and practice their knowledge of pedagogy and reading assessment and instruction."

Reconsidering the Counterfactual

In recent years, a quartet of authors (Lemons, Fuchs, Gilbert, & Fuchs, 2014) completed a secondary analysis of existing data sets from 1990 through 2005 during the evaluation of the implementation of a reading program in a large school district. The authors reported that even though the students who received the intervention during the final study (reported in 2005) experienced the same gains from the intervention as previous samples, the control students had received an even greater acceleration in gains from the standard reading curriculum across the years. As a result, later analyses were unable to show significance, and in some cases, the control-group students actually outperformed the intervention students (as in our study).

Lemons and colleagues (2014) argued that it is important to hold a more nuanced view of the counterfactual (i.e., the outcome performance of the control or comparison group as compared to the intervention group) in light of the major changes initiated in the country following the release of the National Reading Panel (2000) and No Child Left Behind Act. (In fact, Gamse et al., 2001, noted that nearly $6 billion was spent on the Reading First initiative alone.) As instructional practices have become more data-informed and research-based, the results of decades upon decades of previous research have made their way into the standard education experience for many children. As a result, these authors noted that multiple scholars have argued that the research that spurred these changes have actually made it more difficult to achieve effect and significance in modern study by generating great interference through increased positive outcomes from control samples (Hernán & VanderWeele, 2011; Schwartz, Gatto, & Campbell, 2011).

In addition, the authors noted that not only do practices change over time in one place, they also vary at one time in different places (Coyne et al., 2013), which deeply affect school practices, student achievement, and estimates of the value of any one program. This was certainly the case in the evaluation of this program, which was operating state wide in a multiplicity of buildings, where we saw evidence of differing practices. As noted earlier, we also used data on building success on statewide reading measures and state-modeled school report cards for secondary analysis, which provided further evidence of building-to-building variation in practices and success supporting positive student outcomes in reading.

Thus, as a result of Lemons and colleagues' (2015) effort in examining extensive past data sets and the contributions of others (Hernán & VanderWeele, 2011; Schwartz et al., 2011), the literature has suggested—and we tend to agree—that the analytic achievement of significance and effect size should not be the only goal of program evaluation. Rather than just focusing on an intervention group outperforming a control group, one could consider an interpretation around a similarly performing intervention group that takes less time, is simpler, is less

expensive, or is more sustainable than the alternative. With this view in mind, a volunteer program requiring minimal training and producing similar if not significantly different results as compared to another program requiring implementation by a school professional could be demonstrated as a real asset to districts lacking the resources to hire personnel or the time to train them.

Formative Impacts on Programming

As a result of the evaluation experience, the America Reads - Mississippi program looks different in the 2016-2017 school year than it did during the 2014-2015 school year. For example, the administration team strengthened its application materials and further refined and provided more hours for pre-assignment training. At the same time, some of the processes and tools created for the evaluation year have continued to inform administrators about implementation adherence at local levels. These tools include surveys of members, a site observation rubric, and a member observation rubric.

The team has also introduced new tools designed to support the mission of each member at the local level, as well as to promote positive learning experiences for students during sessions. These include a guide for members to use during their weekly meetings with school staff to ensure that they are quality and focus on instruction. In addition, administrators are publishing a monthly electronic newsletter with tutoring tips for members, which includes links to resources. Finally, administrators have distributed some curricular resources to each member for use with students (e.g., phonics flash cards, phonemic awareness sound box kits, alphabet kits for letter recognition, and alphabet letter tiles). Please see Appendix A for more information about the America Reads - Mississippi journey through its three-year evaluation cycle.

Conclusions

In conclusion, this evaluation adds to a growing base of literature related to the implementation of a statewide tutoring program targeting reading supported by volunteer tutors in local school buildings. It is important to note the challenges to implementation adherence encountered and moderated by state administrators in light of findings that support evidence in the literature that the quality of the tutoring curriculum, setting, and support, as well as the professional skills of the tutor, might impact student gains. Further research is needed to investigate volunteer tutor program structures supporting high levels of local control, as with America Reads - Mississippi. In addition, future evaluation teams should reconsider the counterfactual and consider the benefit of low-cost and flexible tutoring programs that deliver any impacts, even when found not to be statistically significant.

REFERENCES

Aberson, C. L. (2011). *Applied power analysis for the behavioral sciences*. Routledge.

Adams, A.-M., & Gathercole, S. E. (2009). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders*.

Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough? *Exceptionality*, *13*(4), 195–208.

Alexander, J. E., & Filler, R. C. (1976). *Attitudes and reading*. Newark, Del.: International Reading Association.

Allington, R. L. (1994). What's special about special programs for children who find learning to read difficult. *Journal of Literacy Research*.

Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research & Practice*, *19*(2), 116–129.

American Academy of Arts & Sciences. (1999). Intergenerational Literacy Tutoring Project. *Bulletin of the American Academy of Arts and Sciences*, *52*(4), 25–57.

Applegate, A. J., & Applegate, M. D. (2004). The Peter Effect: Reading habits and attitudes of preservice teachers. *The Reading Teacher*, *57*(6), 554–563.

Askew, B. J., & Simpson, A. (n.d.). Does one-to-one teaching really matter? Retrieved from http://wwwp.oakland.edu/Assets/Oakland/readingrecovery/files-and-documents/differentiatedinstructionthroughone-to-oneteaching.pdf

Baker, E. (2014). *Exploring meanings of professional development: Teacher perspectives*. University of Central Missouri. Retrieved from https://mospace.umsystem.edu/xmlui/handle/10355/43704

Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, *34*(4), 452–477.

Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, *35*(4).

Braze, D., Katz, L., Magnuson, J. S., Mencl, W. E., Tabor, W., Van Dyke, J. A., … Shankweiler, D. P. (2016). Vocabulary does not complicate the simple view of reading. *Reading and Writing*, *29*(3), 435–451.

Burns, M. K., Senesac, B. J., & Silberglitt, B. (2008). Longitudinal effect of a volunteer tutoring program on reading skills of students identified as at-risk for reading failure: A two-year follow-up study. *Literacy Research and Instruction*, *47*(1), 27–37.

Burns, M. K., Senesac, B. V., & Symington, T. (2003). The effectiveness of the hosts program in improving the reading achievement of children at-risk for reading failure. *Literacy Research and Instruction*, *43*(2), 87–103.

Cawthon, C. H., & Maddox Jr, J. S. (2009). Small-group versus one-on-one educational therapy for struggling readers and writers. *Unpublished Doctoral Dissertation, Barry University*.

Chang, S. H. (2011). Grade level and gender differences in a school-based reading tutoring program. *Reading Horizons*, *51*(1).

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237–248.

Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., … Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, *6*(1), 1–23.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Davenport, S., Arnold, M., & Lassmann, M. (2004). The impact of cross-age tutoring on reading attitudes and reading achievement. *Reading Improvement*, *41*(1), 3.

Dogan, E., Ogut, B., & Kim, Y. Y. (2015). Early Childhood Reading Skills and Proficiency in NAEP Eighth-Grade Reading Assessment. *Applied Measurement in Education*, *28*(3), 187–201.

Dorn, L., & Allen, A. (1995). Helping Low-Achieving First-Grade Readers: A Program Combining Reading Recovery Tutoring and Small-Group Instruction. *ERS Spectrum*, *13*(3), 16–24.

Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*(4), 605.

Elley, W. B. (1992). The Language Background. *International Handbook of Reading Education*, 225.

Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. A. (1996). Assessing motivation to read. *The Reading Teacher*, *49*(7).

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). Reading First Impact Study. Final Report. NCEE 2009-4038. *National Center for Education Evaluation and Regional Assistance*.

Gómez-Bellengé, F. (2004). Over one million served… and counting: National data update 2002–2003. *The Journal of Reading Recovery*, *3*(2), 64–67.

Guthrie, J., McGough, K., & Wigfield, A. (1994). *Measure Reading Activity: An Inventory*. Research Report.

Hamman, D. (n.d.). Effects of being a tutor on adolescents' academic efficacy, and value for school achievement. Retrieved from http://www.just-the-facts.us/uploads/1/9/3/6/19368805/ttips_report_year_2.pdf

Harrison, L. (2002). A study on the complementary effects of Reading Recovery and small group instruction for reversing reading failure. *Research Summary*, (210–3).

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Henk, W. A., Marinak, B. A., & Melnick, S. A. (2012). Measuring the reader self-perceptions of adolescents: Introducing the RSPS2. *Journal of Adolescent & Adult Literacy*, *56*(4), 311–320.

Henk, W. A., & Melnick, S. A. (1992). The initial development of a scale to measure" perception of self as reader." In *Literacy research, theory, and practice: Views from many perspectives. 41st Yearbook of the National Reading Conference* (pp. 111–117).

Hernán, M. A., & VanderWeele, T. J. (2011). Compound treatments and transportability of causal inference. *Epidemiology (Cambridge, Mass.)*, *22*(3), 368.

Hernandez, G. M. (2011). *Parent involvement: perceptions of parents and teachers*. California State University, Sacramento.

Hitchcock, C. H., Prater, M. A., & Dowrick, P. W. (2004). Reading comprehension and fluency: Examining the effects of tutoring and video self-modeling on first-grade students with reading difficulties. *Learning Disability Quarterly*, *27*(2), 89–103.

Institute of Education Sciences. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user-friendly guide*. US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Invernizzi, M., Rosemary, C., Juel, C., & Richards, H. C. (1997). At-risk readers and community volunteers: A 3-year perspective. *Scientific Studies of Reading*, *1*(3), 277–300.

Johnson, E., & Pesky, L. (n.d.). How to develop an effective Tier 2 system. *How to Develop an Effective Tier 2 System*.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly*, *31*(3), 268–289.

Kazelskis, R., Thames, D., & Reeves, C. (2004). The elementary reading attitude survey: Factor invariance across gender and race. *Reading Psychology*, *25*(2), 111–120.

Kazelskis, R., Thames, D., Reeves, C., Flynn, R., Taylor, L., Beard, L. A., & Turnbo, D. (2005). Reliability and stability of Elementary Reading Attitude Survey (ERAS) scores across gender, race, and grade level. *Professional Educator*, *27*, 29–37.

Kershaw, S., & Schatschneider, C. (2012). A latent variable approach to the simple view of reading. *Reading and Writing*, *25*(2), 433–464.

Lazarus, B. D., & Callahan, T. (2000). Attitudes toward reading expressed by elementary school students diagnosed with learning disabilities. *Reading Psychology*, *21*(4), 271–282.

Leal, D., Johanson, G., Huang, C.-C., & Toth, A. (2004). Increasing at-risk students' Literacy skills: Fostering success for children and their preservice Reading endorsement tutors (Reprint from Vol. 41, Spring/2004, Number 1). *Reading Improvement*, *41*(2), 75.

Lee, J., Grigg, W., Donahue, P., & others. (2007). The nation's report card. *Reading*, *496*.

Lee, Y. S., Morrow-Howell, N., Jonson-Reid, M., & McCrary, S. (2010). The effect of the Experience Corps(R) program on student reading outcomes. *Education and Urban Society*, *44*(1), 97–118. https://doi.org/10.1177/0013124510381262

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, *43*(5), 242–252.

Lin, D., & Lin, M. D. (2009). Package "VIF." Retrieved from http://www.piotrkosoft.com/pub/Linux/software/mirrors/CRAN/web/packages/VIF/VIF.pdf

Lonigan, C. J., & Shanahan, T. (2009). Developing Early Literacy: Report of the National Early Literacy Panel. Executive Summary. A Scientific Synthesis of Early Literacy Development and Implications for Intervention. *National Institute for Literacy*.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.

Markovitz, C. E., Hernandez, M. W., Hedberg, E. C., & Silberglitt, B. (2014). Impact Evaluation of the Minnesota Reading Corps K-3 Program. *Corporation for National and Community Service*.

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*(3), 223–262.

Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*.

Mathewson, G. C. (1994). Model of attitude influence upon reading and learning to read.

McKenna, M. C. (1994). Toward a model of reading attitude acquisition. *Fostering the Love of Reading: The Affective Domain in Reading Education*, 18–40.

McKenna, M. C. (2001). Development of reading attitudes. *Literacy and Motivation: Reading Engagement in Individuals and Groups*, 135–158.

McKenna, M. C., Conradi, K., Lawrence, C., Jang, B. G., & Meyer, J. P. (2012). Reading attitudes of middle school students: Results of a US survey. *Reading Research Quarterly*, *47*(3), 283–306.

McKenna, M. C., & Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. *The Reading Teacher*, *43*(9), 626–639.

McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly*, 934–956.

Meier, J. D., & Invernizzi, M. (2001). Book buddies in the Bronx: Testing a model for America Reads. *Journal of Education for Students Placed at Risk*, *6*(4), 319–333.

Morris, D., Shaw, B., & Perney, J. (1990). Helping low readers in grades 2 and 3: An after-school volunteer tutoring program. *The Elementary School Journal*, *91*(2), 133–150.

Morris, D., Tyner, B., & Perney, J. (2000). Early Steps: Replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology*, *92*(4), 681.

Morrow, L. M., & Woo, D. G. (2001). *Tutoring Programs for Struggling Readers: The America Reads Challenge. Rutgers Invitational Symposia on Education.* ERIC.

National Center for Education, & Statistics. (2013). *The Nation's Report Card: A First Look: 2013 Mathematics and Reading (NCES 2014-451)*. Washington, D.C.: Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf

Navarra, E. M. (2011). *Student Attitudes Toward and Perceptions of Reading in a First Grade Classroom*. The College at Brockport: State University of New York, Education and Human Development.

Nimon, K. (2010). Regression commonality analysis: Demonstration of an SPSS solution. *Multiple Linear Regression Viewpoints*, *36*(1), 10–17.

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*(2), 457–466.

Nimon, K., & Oswald, F. L. (2013). Understanding the results of multiple linear regression beyond standardized regression coefficients. *Organizational Research Methods*, *16*(4), 650–674.

O'Connor, R. E. (1997, February). *Layers of intervention*. Presented at the Pacific Coast Research Conference, La Jolla, CA.

Osborn, J., Freeman, A., Burley, M., Wilson, R., Jones, E., & Rychener, S. (2007). Effect of tutoring on reading achievement for students with cognitive disabilities, specific learning disabilities, and students receiving Title I services. *Education and Training in Developmental Disabilities*, 467–474.

Parker, C. T. (2004). *An evaluation of student reading attitudes: does ability affect attitude?* University of North Carolina at Wilmington.

Pena, N. A. P. (2008). *The power of reading: A multilevel study of the longitudinal effect of a paired intergenerational reading aloud program on academically at-risk elementary students' reading attitudes, reading motivation and academic achievement*. ProQuest.

Ritter, G., Albin, G., Barnett, J., Blankenship, V., & Denny, G. (2006). The effectiveness of volunteer tutoring programs: A systematic review. *Campbell Systematic Reviews*, *2*(7).

Rodrigues, V. (2013). Publication and reporting biases and how they impact publication of research. *Editage Insights (29-10-2013)*.

Rosenbaum, P., & Rubin, D. (1983). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38.

Roth, D., Worrell, F., & Gabelko, N. (2002). Elementary reading attitude survey scores in academically talented students. In *annual meeting of the American Educational Research Association, New Orleans, LA*.

Schmitt, E. (2009). *Four case studies: The reading attitudes and practices of teachers and students In second grade*. Bowling Green State University. Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=bgsu1237813672&disposition=attach ment

Schwartz, S., Gatto, N., & Campbell, U. (2011). What would have been is not what would be: counterfactuals of the past and potential outcomes of the future. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*.

Seitz, L. (2010). Student attitudes toward reading: A case study. *Journal of Inquiry and Action in Education*, *3*(2), 3.

Shanahan, T. (1998). On the effectiveness and limitations of tutoring in reading. *Review of Research in Education*, *23*, 217–234.

Snow, C. E., Burns, M. S., Griffin, P., & others. (1998). *Preventing reading difficulties in young children*. National Academies Press.

Stanovich, K. E. (1986). Cognitive processes and the reading problems of learning disabled children: Evaluating the assumption of specificity. *Psychological and Educational Perspectives on Learning Disabilities*, 87–131.

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1998). Literacy experiences and the shaping of cognition. *Global Prospects for Education: Development, Culture, and Schooling*, 253–288.

The Education Trust—West Response to the 2009 Nation's Report Card. (2010, March 24). Retrieved November 1, 2016, from https://west.edtrust.org/press_release/the-education-trustwest-response-to-the-2009-nations-report-card-2/

Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*(4), 383–401.

Torgesen, J., Wagner, R., Rashotte, C., Lindamood, P., Rose, E., & Conway, T. (1999). Preventing Reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, *91*(4), 579–593.

Tran, L., Sanchez, T., Arellano, B., & Swanson, H. L. (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities*, *44*(3), 283–295.

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *The Elementary School Journal*, *110*(2), 228–245.

U. S. Department of Education. (1979). School volunteer development project. *Proposal Submitted to the Program Effectiveness Panel of the National Diffusion Network: Washington, DC*.

Unrau, N., & Schlackman, J. (2006). Motivation and its relationship with reading achievement in an urban middle school. *The Journal of Educational Research*, *100*(2), 81–101.

U. S. Department of Health and Human Services. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for*

reading instruction. National Institute of Child Health and Human Development, National Institutes of Health.

U.S. Department of Education. (2000). *Twenty-second annual report to Congress on the implementation of the Individuals with Disabilities Education Act* (p. Washington, DC: Government Printing Office). Retrieved from https://www2.ed.gov/about/reports/annual/osep/2001/toc-execsum.pdf

Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly*, *20*(2), 126. https://doi.org/10.2307/1511219

Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities*, *33*(6), 579–590.

Vadasy, P. F., & Sanders, E. A. (2008). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology*, *100*(2), 272.

Vandervelden, M. C., & Siegel, L. S. (1997). Teaching phonological processing skills in early literacy: A developmental approach. *Learning Disability Quarterly*, *20*(2), 63–81.

Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009). Achievement gaps: How black and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress. Statistical Analysis Report. NCES 2009-455. *National Center for Education Statistics*.

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education*, *24*(5), 301–315.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, *88*(4), 601.

Venables, W., & Ripley, B. (2002). Data Manipulation. In *Modern Applied Statistics with S* (pp. 13–39). Springer.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2015). Meta-analyses of the effects of Tier 2 type Reading interventions in grades K-3. *Educational Psychology Review*, *28*(3), 551–576. https://doi.org/10.1007/s10648-015-9321-7

Wasik, B. A. (1998). Volunteer tutoring programs in Reading: A review. *Reading Research Quarterly*, *33*(3).

Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 179–200.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *Journal of Educational Psychology*, *89*(3), 420.

Worrell, F. C., Roth, D. A., & Gabelko, N. H. (2006). Elementary reading attitude survey (ERAS) scores in academically talented students. *Roeper Review*, *29*(2), 119–124.

APPENDIX A. AMERICA READS - MISSISSIPPI PROGRAMMING ACROSS A THREE-YEAR EVALUATION CYCLE

This report documents findings from a 2015-2016 Impact Evaluation of the America Reads - Mississippi program. However, it is important to note that this school year was actually the second of a three-year evaluation process. From a programming standpoint, both year 1 and year 3 were of equal importance to the school year reported in this document.

During the first year, the program's central administration team established valuable relationships with districts and schools who returned data use agreements for the upcoming year. The team then constructed the supports and parameters necessary for hiring an evaluation consultant.

During year 3, the 2016-2017 school year tutoring began with the program continuing processes and practices initiated during the second school year. These are:

1. AmeriCorps Member Selection: Developed an Online Applicant Survey that would be completed by applicants slated for recommendation by program and school staff to move to the next level of the selection process. Completion and responses would assist with final selection.

2. 360 Degree Member Support: Success team meetings held at partnering schools with ARM regional staff, AmeriCorps members, classroom teachers, and site supervisors to help clarify the role of ARM members and foster successful relationships between teachers and members. The program director also presented to the MDE Literacy Coaches to garner regular mentoring, coaching, and resourcing for members at the school sites.

3. Training: More extensive tutor training (extended to 2 days) to add sessions dedicated to key foundational skills (preceding the existing full-day MDE training on the 5 components of reading instruction) including classroom management, tutoring expectations (e.g., record keeping, lesson planning, available resources, initial steps, student attitude survey), and America Learns system use training. On the second day of training, MDE literacy coordinators also provided members with additional resources including explicit phonological awareness lesson plans, graphic organizers for vocabulary and comprehension instruction, and a phonics screener that members could administer to students before beginning phonics instruction.

4. Curriculum: Purchase and use demonstration of additional tutoring resources for members to use resources also provided by MDE (e.g., phonics flash cards, phonemic awareness sound box kits, alphabet kits for letter recognition, alphabet letter tiles).

5. Observation: Statewide use of the member observation rubric (formerly used only in the USM region) and continued use of the tutoring fidelity rubric at partnering schools to assess the quality of tutoring environments provided and maintained by schools.

6. Communication: Monthly electronic newsletter (Tutoring Tips) for members that includes links to resources for tutoring (on ARM website) and scripted lessons.

7. Resources for local teams: Development of an agenda/guide for members to use during their weekly meetings to ensure they are quality, instructional meetings and to encourage participation from instructional personnel.

For reference, the entire three -year plan is laid out in Figure A1 from September of 2014 and continuing through September of 2017.

Figure A1. The America Reads - Mississippi three-year evaluation cycle

ARM Evaluation Planning Timeline Year 1 (2014-2015)

September   July
• Develop RFP and hire evaluator • Formalize data collection instruments: STAR REPORTS • Determine data resources, data collection methods, sample size requirements, analytic approach • (1st face to face meeting w/ evaluator)

Beginning Year 2 (2015-2016) New Partners: *TerraLuna*

August	September	October December
• Member and Site Supervisor Training • Attain IRB approval for impact study • Begin to collect STAR scores and assign students to members • Initiate formal roster development	• Members begin tutoring • Members begin using America Learns to report activity	• Members continue tutoring • Members continue to report activity • Student attitude survey and analysis • Collect STAR scores • Formal roster data quality evaluation

Beginning Year 3 (2016-2017)

January March	April May	June August	September December
• Members continue tutoring • Members continue to report activity • Members & Parents survey dissemination • Collect FINAL STAR scores • Formal roster data quality evaluation • DELIVERED: Student attitude survey report	• End of experimental conditions, members may support all students • DELIVERED: Parents & Member Survey Reports	• Members end tutoring • Members end activity reporting **TERRALUNA ANALYZES ALL DATA**	• 2016-2017 members begin tutoring • 2016-2017 members begin reporting activity • DELIVERED: Final Impact Evaluation Report • ARM Re-Compete Appl. Due MCVS 11/18/16

APPENDIX B. TABLE OF REVIEWED LITERATURE

Authors provide this table in an effort to provide an accessible synthesis of the research reviewed in the introduction section of this report. All of the studies presented here documented programs that were similar to America Reads - Mississippi. In Table B1, we present the year, authors, and title of the article, as well as details about the intervention design and key findings.

Table B1. Table of reviewed literature

[1]This article is not included the review of literature within the introduction

Year	Authors	Title	Intervention Design	Key Findings
2005	Al Otaiba, S., Schatschneider, C., & Silverman, E.	Tutor-assisted intensive learning strategies in kindergarten: How much is enough?	**Program Studied:** Tutor Assisted Intensive Learning Strategies (TAILS) **Student Population:** Kindergarteners **Group Size:** (Not specified, likely 1:1 based on number of students & tutors) **Setting:** In an unused classroom or, in one school, the lunchroom before and after lunchtime. **Curriculum:** Based on scientifically based reading research, adapted to be consistent with classroom lessons. **Tutor Background:** Community members paid $10 per hour **Tutor Training:** 13 hours over 3 sessions (2 before the intervention and one the week after) **Intervention Length:** Tutoring sessions were 30 minutes **Treatment Groups:** Students at four schools were randomly assigned to one of three groups: a) tutoring four days a week, (b) tutoring two days a week, or (c) a control condition that provided small-group storybook reading two days a week.	-Students in who were tutored four days a week outperformed students in either the two-day or control groups on the three reading measures used with effect sizes ranging from .79 to .90. -Students in both the two-day and four-day groups outperformed students in the control group on "Blending Sounds" by a significant margin (ES = 0.68). -The program effects were achieved despite limitations such as no funding to hire a reading specialist who could tailor lessons for individual students (a key part of other programs), schools being challenged to protect the time for tutoring beyond the 90-min literacy block, and having a designated quiet area for tutoring.
2000	Baker, S., Gersten, R., & Keating, T.	When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training	**Program Studied:** Start Making a Reader Today (SMART) program **Student Population:** First and second grade students having "trouble learning the basics" **Group Size:** 1:1 **Curriculum:** There is no curriculum but rather a broad framework (30 minutes of training). **Tutor Background:** Adult Volunteers **Tutor Training:** 1-2 hours **Intervention Length:** 30 minute sessions two times per week that took place over two years	-The group receiving the intervention made greater growth on word identification than their grade-level peers who did not receive the intervention. -At the end of 2nd grade, the treatment group also had higher scores on reading fluency and word comprehension. -Tutors frequently expressed a desire for more guidance (in lieu of formal training hours) on how to help specific students. -Roughly half of tutors did not return to participate in a second year.
2010	Gattis, M. N., Morrow-Howell, N., McCrary, S., Lee, M., Jonson-Reid, M., McCoy, H., Tamar, K., Molina, A., & Invernizzi, M.[1]	Examining the Effects of the New York Experience Corps Program on Young Readers	**Program Studied:** NY Experience Corps **Student Population:** Teachers identify students in need. **Group Size:** (Not specified, likely 1:1 based on number of students & tutors) **Setting:** One-on-use in a separate space provided by the school. **Curriculum:** Book Buddies **Tutor Background:** Community members aged 55 and over supported by 19 paid personnel, each assigned to administer the program at one school. Many staff are former AmeriCorps members. **Tutor Training:** 32 hours in initial two-week classroom training followed by 16 hours of "on-the-job training." **Intervention Length:** Tutoring sessions were 45 minutes. On average each student received 48 sessions. **Treatment Groups:** 288 identified students were randomly assigned to either the intervention or control group.	-The intervention group outperformed the control group in pre-test to post-test change measures controlling for within-class and within-school differences. Effect sizes were 0.45 on one measure and 0.20 on another measure. -The study offers evidence that programs can use older adults to support positive outcomes for students. These individuals are growing in number, are looking to contribute in their communities, and are more reliable, according to the author.

| 2007 | Osborn, J., Freeman, A., Burley, M., Wilson, R., Jones, E., & Rychener, S.[1] | Effect of tutoring on reading achievement for students with cognitive disabilities, specific learning disabilities, and students receiving Title I services | **Program Studied:** Project MORE, which implements two individualized reading interventions: HOSTS (Help One Student To Succeed) and the Reading-Tutors Program

Student Population: Students with cognitive disabilities, specific learning disabilities, and students receiving Title I services

Group Size: 1:1

Setting: Typically occurred in the classroom but sometimes in a hallway

Curriculum: 400 comprehensive lesson plans and assessment tools patterned after DIBELS.

Tutor Background: Community volunteers

Tutor Training: Two hours prior to mentoring and then regular feedback from a coordinator

Intervention Length: 30 minutes per session, 3-4 days per week | -Students who participated in tutoring had higher month-for-months gains and had higher reading scores than students in the comparison group over a 6-month period.

- Title 1 students and students with specific learning disabilities had statistically significant higher levels on DIBELS and WJIII than their peers who didn't receive the intervention. Those students with cognitive disabilities were significantly higher on DIBELS. |
| 2010 | Lee, Y. S., Morrow-Howell, N., Jonson-Reid, M., & McCrary, S. | The effect of the Experience Corps® program on student reading outcomes. | **Program Studied:** Experience Corps® (EC)

Student Population: Elementary students who are struggling readers

Group Size: 1:1

Curriculum: Chosen by site (some included Book Buddies, Reading Coaches, and the Brigance Inventory of Basic Skills)

Tutor Background: Older adult volunteers

Tutor Training: 15-32 hours

Intervention Length: Two-to-four sessions per week for 30-40 minutes per session, averaging between 35 and 58 sessions for each of the three locations included in the study. | - Findings indicated that intervention students made statistically greater gains over the academic year on passage comprehension and grade-specific reading skills, especially those who received 35 tutoring sessions or more.

-Students participating in this study were mostly ethnic minorities and from families of low socioeconomic status and thus at higher risk of reading difficulty. The findings from this study suggest that this program effectively reaches vulnerable students and offers an intervention that can reduce reading disparities. |

2012	Reutzel, D. R., Petscher, Y., & Spichtig, A. N.[1]	Exploring the value added of a guided, silent reading intervention: Effects on struggling third-grade readers' achievement.	**Program Studied:** Computer-based Reading Plus silent reading program **Student Population:** Struggling third-grade readers who were retained at grade level as a result of poor performance on the reading portion of a state assessment. **Group Size:** Student and computer **Setting:** Computer in a classroom **Curriculum:** Computer-based, guided silent reading fluency program using a combination of reading passages and comprehension questions **Tutor Background:** Computer-based **Intervention Length:** Additional 30 minutes of reading each day over one school year. **Treatment Groups:** The treatment group received 30 minutes of computer-guided silent reading on Reading Plus. The control group received an additional 30 minutes with a guided classroom lesson using Soar to Success, Essential Elements of Reading: Vocabulary, Voyager Passport, or Earobics.	-The silent guided reading intervention had an effect size 1 full standard deviation when looking the comparison groups' posttest scores, indicating that this type of supplementary instruction can help struggling third grade readers develop core reading competencies. -After the intervention, the majority of the students were making progress to move to the next grade level.

1997	Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E.	The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers	**Program Studied:** **Student Population:** First graders **Group Size:** 1:1 **Setting:** Unused room in the school. **Curriculum:** 100 30-minute lessons with explicit scripts for tutors to follow in teaching each skill. Lessons included as many tasks as possible that had been identified previous studies to contribute to gains in phonological awareness. **Tutor Background:** Community volunteer **Tutor Training:** Initially trained for six hours, followed by three hours of training once they had experience in the field. **Intervention Length:** Tutoring began in November and took place after school. The intervention ended in May at a maximum of 23 weeks and 53 hours of instruction	-Tutored students' post-test means exceeded those for the control group on all measures; effect sizes ranged from .15 to .56. -Volunteer tutors who closely followed the lessons were able to achieve significant effect sizes in all of the early reading skill areas that were assessed. -The research team learned that when recruiting volunteer tutors, it is critical to select tutors who are dedicated to helping young students and who are able to learn how to conduct lessons with consistency and care. -The research team also learned that tutors and students needed regular monitoring for successful program implementation.
2000	Vadasy, P. F., Jenkins, J. R., & Pool, K.	Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities.	**Program Studied:** **Student Population:** First graders at risk for a learning disability. **Group Size:** 1:1 **Setting:** Unspecified **Curriculum:** The authors developed 100 scripted lessons that targeted decoding, rime analysis, and story reading. **Tutor Background:** Non-professional tutors who were paid $5 an hour Tutor Training: received 14 hours of training, **Intervention Length:** 30 minutes, 4 days a week, for an entire school year.	-At year end, tutored students significantly outperformed students in the control group on measures of reading, spelling, and decoding with effect sizes ranging from .42 to 1.24. -The nonprofessional tutors more precisely implemented the tutoring program and exhibited better teaching skills when they had regular support from a supervisor.

APPENDIX C. DEVELOPING SCALES FROM EXISTING INSTRUMENTS

In order to better understand the predictive power of member- and school-level factors, a more nuanced insight of the instruments was necessary. To better discern the explanatory power of the instruments in our secondary analyses, we conducted a principal component analysis (PCA) to determine how individual items loaded on factors. We employed the use of PCA, a form of exploratory analyses, rather than a confirmatory technique, because of the untested nature of the instruments. As a result, we were able to identify sub-scales within each of two main instruments: 1) the administrator evaluations of member quality, and 2) the end-of-year member settings self-report.

Examining the Administrator Evaluation of Member Quality Instrument for Inclusion in Analysis

Our analysis of the administrator evaluation of member quality instruments included the following activities:

1. Examining the reasonableness of averaging the two time-period measures to maximize the stability of observations,
2. Examining the reliability of the measure overall,
3. Examining the reliability of the measure minus a final question that served as a check on "overall" feelings,
4. Conducting exploratory factor analyses (PCA) to determine how items might best load depending on potential multi-collinearity,
5. Confirming the resulting factors based on their correlations, and
6. Including identified factors in the predictive models.

The particular evaluation instrument that we investigated included a standard agreement scale that ranged from "strongly disagree" to "strongly agree." The following prompts after the "overall" question had been deleted:

1. Attendance during the school day
2. Attendance in after-school programs
3. Attendance at program-required member training
4. Participation in community service projects (i.e., planning, attending, etc.)
5. Completion of program paperwork in a timely, neat, and accurate manner
6. Submission of timesheets on America Learns in a timely and accurate manner

7. Evidence of understanding ARM program objectives and shows commitment (we understand this prompt to be double-barreled, however we chose to include in our analysis)
8. Demonstration of the AmeriCorps principle of "Getting Things Done"

Averaging Across Administrations

As a result of accessing data for both mid-year and end-of-year evaluations we questioned whether ratings were stable enough that it would be reasonable to use a mean for each member in later calculations. As a result, we attempted to determine whether there was any consistency in the direction of the difference. Results indicated that the mean difference ranged approximately +/- 0.25 with the highest frequency of differences occurring with Qs 6 and 7, followed by Q3 and Q4. Table C1 shows a sample of the relevant descriptive statistics that were examined during the comparison.

Table C1. Descriptive Data Resulting From Survey Analysis

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Mean difference	-0.02	0.05	0.19	0.15	-0.02	-0.05	0.27	0.20	0.12
Std. Error of Mean	0.11	0.16	0.06	0.06	0.09	0.09	0.07	0.07	0.07
Median	0	0	0	0	0	0	0	0	0
Mode	0	0	0	0	0	0	0	0	0
Std. Deviation	0.88	0.97	0.47	0.51	0.69	0.73	0.58	0.57	0.52
Range	3	4	2	3	3	5	2	2	2
Minimum	-2	-2	-1	-1	-2	-3	-1	-1	-1
Maximum	1	2	1	2	1	2	1	1	1
Sum	**1**	**2**	**12**	**9**	**1**	**3**	**17**	**12**	**7**

As a result of the analysis, we determined that averaging the two ratings could serve to increase the stability of each indicator and we proceeded with using each.

Reliability

As a result of internal consistency calculations (Cronbach's Alpha), all nine items had a standardized alpha of 0.825. However, we suspected that the 9th question over-inflated the reliability as it served as a proxy for the overall rating of questions 1-8. As a result, we completed three different sets to confirm our suspicions. The first sought to determine whether the full scale was reliable as a whole. Recognizing the sample- and construct-dependency associated with Cronbach's alpha, we reduced the item set to exclude the overall question.

While this addressed one problem, it did not address the issue of combining multiple factors in a single reliability analysis. Thus, we calculated Cronbach's alpha on the final 5-item scale that exhibited an adequate standardized alpha reliability of 0.815. The results are displayed in Table C2 below.

Table C2. Internal Consistency Calculations for Distinct Sets of Evaluation Data

Scale	Standardized Alpha	Number Of Items Used
Full scale	0.825	9
Reduced scale (removing Q9)	0.764	8
Factor 1 (based on 3-factor solution)	0.815	5

We believed that reliability testing on the remaining 1-item and 2-item scales, the remaining factors, was unnecessary. Given our analysis, we argue that the resulting 5-item scale is *sufficiently* reliable for use as is. In the future, it would be more appropriately used as a single subscale based on the items that loaded into Factor 1.

Programmer Note. For future use, the program should employ sound psychometric techniques to develop the scale, which could potentially include the following:

1. Begin with a qualitative analysis of themes deemed as critical to program success through focus group or interview methods,
2. Allow for natural themes to emerge from which items could be developed based on thematic analysis,
3. Field test potential items to determine which ones should be maintained as part of a potential factor or subscale of member success,
4. Confirm the value to understanding the program and the literature by including this in studies using structural equation modeling in program evaluation and analysis.

Principal Components Analysis and Examining Correlations

For the scale itself, we completed iterations of PCA given the marginal reliability exhibited by the scale. We considered these efforts exploratory given that (1) there was not a formalized development plan to inform item development, (2) there was no training providing to the raters, potentially injecting rater noise into the observations, (3) there were no clear emergent relationships among the variables, and (4) unexpected patterns that didn't connect with what the face grouping could be. Given these considerations, we implemented 3 rounds of PCA:

1. Exploratory PCA with no rotation to determine what base Eigenvalues and factor loadings would be.
2. Exploratory PCA with orthogonal rotation (i.e., Varimax) to determine whether there was any improvement with the assumption that the factors were independent (i.e., assuming no multicollinearity).
3. Exploratory PCA with oblique rotation (i.e., Direct Oblimin) to determine whether there was any improvement with the assumption that the factors were not independent (i.e., no multicollinearity).

In all cases, 3 factors emerged, with the final factor loading structure being:

- Factor 1 = Q1, Q5, Q6, Q7, and Q8
- Factor 2 = Q3
- Factor 3 = Q4 and Q2

Based on various independent correlations between the resulting factors, we checked for the feasibility of a two-factor structure (Factor 2 being only Q3), which was unreasonable given the performance of the individual items. The Eigenvalues suggested at least three factors were present, but not conclusively. In other words, the three main factors that emerged only accounted for 80% of the variance in the data, though it may also be the case that they were collinear.

After determining the possible factor structure, correlations among factors were conducted to confirm researcher assumptions. Ultimately, the first (5-item) factor is correlated to Factors 2 and 3, but Factor 2 and 3 are not correlated.

Conclusion

After running through examinations of the items, our recommendation was to include the three factors (average of 5 items, a single item of Q3, and average of Q4 and Q2) in the model, though we expected that Factor 1 would be the most useful as a predictor. As noted in the results, Factor 3 was not a significant predictor in any model. The two surviving factors (and corresponding items) were named member quality and member attendance.

As noted previously, the major takeaway for future evaluation efforts stems from developing a sound instrument to help inform program improvement and informing interpretations of the outcome. This will require some investment of time and resources, but could be very beneficial in the end.

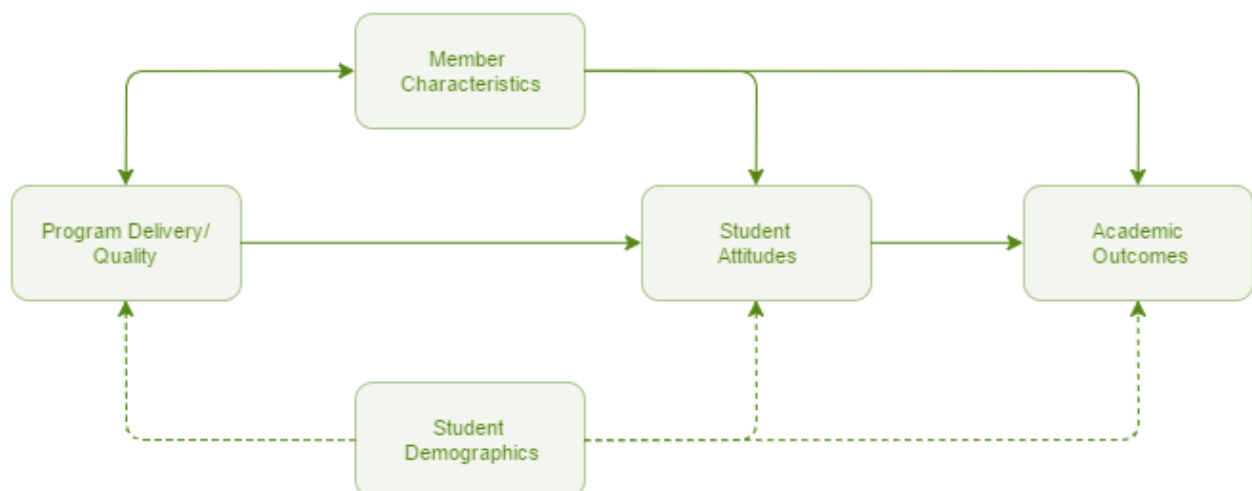# APPENDIX D. SECONDARY ANALYSIS: COMMONALITY COEFFICIENT

Appendix D presents the full results of the Commonality Coefficient analyses referenced in the narrative of the report (Nimon et al., 2008; Nimon & Oswald, 2013). As a reminder, we utilized these techniques as we sought to better understand the relationship among the variables included in the study and the potential impact they may have had on student academic or attitudinal outcomes.

Background and Interpretation

The following analysis allows us to better understand a predictor's contribution to an outcome of interest in isolation and combination with other predictors (Nimon, 2010). In other words, this analysis allows us to identify the areas that program staff should focus on to yield the most significant improvements in student outcomes given a limited set of resources. The following findings are presented to attempt to conceptualize the relative value of future efforts and how that relates to previous findings.

Before examining the way in which the predictors contributed to the outcomes, it is important to first name the possible theory of action behind the program that inform our analyses. It is expected that there are many characteristics of the program, students, and environment that contribute to student performance on an outcome like the STAR ELA or Early Literacy assessment. However, it is important to identify which characteristics, or variables, might be intermediate steps on the way to the student outcomes. A high-level theory of action is presented below in Figure D1:

Figure D1. Theory of action connecting programming to student outcomes

The proposed high-level theory of action is intended to help contextualize the CA coefficients and percentage of variance explained by the unique and combined predictors. As a point of clarification, not every predictor was included in the CA because every predictor increases the complexity of the model and its results exponentially. For example, if 2 variables are used, then a possible 3 results exist (V1, V2, and V1*V2). If 3 variables are used, then a possible 6 results exist (V1, V2, V3, V1*V2, V1*V3, V2*V3). Continuing this, 4 variables result in 24 possible combinations, 5 result in 120, and so on. Therefore, we strategically selected those variables that were the strongest predictors in the previous models using these steps: 1) Select all independent variables that are significant predictors at the .05 level or better, then 2) Conduct the analyses with five variables for CA if the output permits.

Results

The CA results can be conceptualized as being done in two separate parts. As an initial set of analyses, we examined the unique and combined predictive power of certain variables to student attitudes. In a secondary set of analyses, we examined the unique and combined predictive power of certain variables to academic outcomes. The two sets of examinations are presented below.

Student Attitudes. For the first set of analyses, various combinations of variables were included to determine the unique and shared ability to predict student attitudes toward academics and toward recreational reading. These are described in further detail below.

As noted in the previous section, the linear model using Academic Attitudes as an outcome accounted for approximately 13% (R^2 = .1292, p<.0001) of the variability in the outcome data. Based on an examination of the linear model, we included the following variables based on their significance in the model:

- *Member quality*
- *Member attendance*
- *Gender*
- *Number of weeks*
- *Total evaluation intervention minutes*
- *Observed adherence*

Using only these variables, it is important to note that this model only accounted for approximately 7% (Adjusted R^2 = .0705, p <.0001) of the variability in student attitudes toward academics, a difference of approximately 6%. While the previous model's additional explanatory power can be attributed to other non-significant predictors, CA results for this

model should be treated with caution as the overall predictive power is lower and may not paint a complete picture of what may affect student academic attitudes. The coefficients and relative percentage of the variance explained for key combinations are presented in Table D1.

Table D1. Commonality Analysis Coefficients for Academic Attitudes

Variable	Coefficient	Percent For Each Variable*
Unique to Member Quality	0.002	3%
Unique to Member Attendance	0.008	10%
Unique to Gender	0.009	12%
Unique to Number of Weeks	0.008	10%
Unique to Total Intervention Minutes	0.043	57%
Unique to Observed Adherence	0.019	26%
Common to Member Attendance, Total Intervention Minutes	0.009	12%
Common to Number of Weeks, Total Intervention Minutes	-0.007	-9%
Common to Total Evaluation Minutes, Observed Adherence	-0.014	-18%
Common to Member Quality, Member Attendance, Total Intervention Minutes	-0.002	-2%

* Totals will not equal 100% as they do not reflect all percentages determined by the commonality analysis.

As seen in the table above, the unique factors were the most powerful predictors in the model with total intervention minutes accounting for the most explanatory power. Furthermore, member attendance and total intervention minutes were also strong relative predictors in combination. Interestingly, member quality was a relatively weak predictor when combined with other variables, and in some cases demonstrated a negative effect when combined with other variables. Finally, while it may be difficult to parse the impact of timing (e.g., number of weeks and total intervention minutes), it might be inferred that the number of weeks may only serve as a proxy measure of program dosage and a more precise indicator of time in program should be used.

Based on the results of the CA, it is recommended that program staff focus again screening efforts for members as member attendance appears to have a compounding effect on student attitudes toward academics, as well as on academic outcomes. Additionally, total intervention minutes appears to be a relative strong predictor, implying that more time under intervention would yield more favorable student attitudes toward academic reading (which may not be surprising to the reader). Furthermore, as suggested by these data, members who are screened

to exhibit higher ratings of quality may enable a more efficient delivery of support to students while they are exposed to tutoring.

As noted in the previous section, the linear model using recreational attitudes as an outcome accounted for approximately 14% (R2 = .1424, p<.0001) of the variability in the outcome data. Based on an examination of the linear model, we included the following variables based on their significance in the model:

- *Gender*
- *Member discipline status*
- *Reported issues tutoring*
- *Total evaluation intervention minutes*
- *Observed adherence*

Using only these variables, it is important to note that this model only accounted for approximately 7% (Adjusted R^2 = .0705, p <.0001) of the variability in student attitudes toward academics, a reduction by half. While the previous model's additional explanatory power can be attributed to other non-significant predictors, CA results for this model should be treated with caution as the overall predictive power is lower and may not paint a complete picture of what may affect student recreational reading attitudes. The coefficients and relative percentage of the variance explained for key combinations are presented in Table D2.

Table D2. Commonality Analysis Coefficients for Recreational Attitudes

Variable	Coefficient	Percent For Each Variable*
Unique to Gender	0.007	9%
Unique to Member Discipline Status	0.001	1%
Unique to Reported Issues Tutoring	0.016	20%
Unique to Total Intervention Minutes	0.036	46%
Unique to Observed Adherence	0.021	27%
Common to Reported Issues Tutoring, Total Intervention Minutes	0.015	19%
Common to Total Intervention Minutes, Observed Adherence	-0.013	-17%
Common to Reported Issues Tutoring, Observed Adherence	-0.006	-7%

* Totals will not equal 100% as they do not reflect all percentages determined by the commonality analysis.

As seen in the table above, only some of the unique factors were the most powerful predictors in the model with total intervention minutes accounting for the most explanatory power, followed by the reported issues tutoring and observed adherence. While these uniquely positively predicted student recreational reading attitudes, combinations that included these three variables had potentially negative impacts on student attitudes. It appears that when observed adherence was included in combination with either the reported issues tutoring or total intervention minutes, attitudes were negatively affected. These findings do not necessarily make logical sense and signal the need to reevaluate all instruments and tools used for data collection in the future.

Despite these findings, program staff should be aware of the potentially distant impact the program may have on recreational attitudes toward reading, as that may be influenced significantly by factors outside the school.

Given the findings from both the academic and attitudinal commonality analyses, program staff should consider where to best focus their resources. Given the potential mediating nature of student attitudes, it may be wise to leverage those variables that appear to be contributing to the highest proportion of variation in the data. These include either member characteristics or program dosage. While student characteristics were expectedly a major predictor, focusing efforts on ensuring the quality of tutors will likely have a large impact on student outcomes for an investment on the planning side of the program. From an implementation standpoint, more precise measures of member quality, continued tracking of intervention dosage, along with other characteristics of the program can aid in program monitoring and ongoing improvement.

Student Achievement Outcomes. For the second set of analyses, various combinations of variables were included to determine the unique and shared ability to predict student results on the Early Literacy assessment and the STAR Reading assessment. These are described in further detail below.

As noted in the previous section, the linear model using Early Literacy as an outcome accounted for approximately 34% (R^2 = .3377, p<.0001) of the variability in the outcome data. Based on an examination of the linear model, we included the following variables based on their significance in the model:

- *Observed adherence*
- *Age range*
- *Age offset*
- *Age in months*
- *Member quality*

Using only these variables, it is important to note that this model only accounted for approximately 32% (Adjusted R^2 = .3186, p <.0001) of the variability in the early literacy data, a difference of approximately 2%. While the previous model's additional explanatory power can be attributed to other non-significant predictors, CA results for this model provides us with a better understanding of what areas program staff may want to focus on in future years. The coefficients and relative percentage of the variance explained for key combinations are presented in Table D3.

Table D3. Commonality Analysis Coefficients for Early Literacy

Variable	Coefficient	Percent For Each Variable*
Unique to Observed Adherence	0.019	5%
Unique to Age Range	0.058	16%
Unique to Age Offset	0.129	35%
Unique to Age in Months	0.270	72%
Unique to Member Quality	0.069	19%
Common to Age Range, Age Offset	-0.033	-9%
Common to Observed Adherence, Age in Months	0.053	14%
Common to Age Offset, Age in Months	-0.125	-33%
Common to Age Offset, Member Quality Factor	-0.031	-8%
Common to Age in Months, Member Quality Factor	-0.037	-10%
Common to Age Range, Age Offset, Age in Months	0.033	9%
Common to Age Range, Age in Months, Member Quality	0.014	4%
Common to Age Offset, Age in Months, Member Quality	0.028	8%
Common to Observed Adherence, Member Quality	-0.008	-2%

* Totals will not equal 100% as they do not reflect all percentages determined by the commonality analysis.

As seen in the table above, the unique factors were the most powerful predictors in the model. However, the inclusion of age in months, age range, and/or age offset (i.e., the farther away the student is from a given age) in combination with each other or other variables plays a strong role in the model. The variables of age range and age offset work with each other to define a student's age relative to peers (i.e., so that a 75-month old in Kindergarten, and a 75-month old are not represented by an age variable predicting identical outcomes). This mix of age variables indicates, as expected, that older students are underperforming, likely due to reasons that would be associated with older students in a given grade (e.g., being held back, being enrolled late, etc.). While these variables cannot be directly controlled by program staff, their impact highlights the need for remediating older students as soon as possible.

The most practically significant predictor that can potentially be directly addressed by program staff are those of member quality and observed adherence. However, in isolation, these two variables are not as powerful as the student characteristics variables. Somewhat surprisingly, it does appear that member quality and observed adherence may mitigate some of the impact of some age variables. This is in contrast with the negative impact that member quality and observed adherence may have when combined with each other. While it is difficult to determine whether this is an artifact of the variables themselves or the tools used to measure these constructs, it appears that across both academic outcomes, member quality, and member attendance may be a prioritized area of focus for the members. As a reminder, member quality includes administrator evaluation of a member's performance including attendance, completion of paperwork, submission of timesheets, understanding objectives, commitment, and demonstration of "getting things done."

Similar to member attendance, this variable may represent intrinsic characteristics of the members. However, program staff may want to improve the screening of potential members to maximize the rating. Furthermore, program staff should consider revisiting the measurement of member quality as described in Appendix C.

Observed adherence appeared to play a small role in the model, both in isolation and in conjunction with other models. It may be possible that more observations using a more refined measure and more stringent evaluation training for those who monitor the members may yield more significant findings in the future.

As noted in the previous section, the linear model accounted for approximately 13% (R^2 = .1285, p<.0001) of the variability in the outcome data for the STAR Reading assessment. Based on an examination of the linear model, we included the following variables based on their significance in the model:

- *Member attendance*
- *Pre-test achievement group*
- *Age range*
- *Student recreational reading attitudes*
- *3rd grade proficiency rate*

Using only these variables, it is important to note that this model only accounted for approximately 12% (Adjusted R^2 = .1194, p <.0001) of the variability in the STAR data, a difference of approximately 1%. While the previous model's additional explanatory power can be attributed to other non-significant predictors, CA results for this model provides us with a better understanding of what areas program staff may want to focus on in future years. The

coefficients and relative percentage of the variance explained for key combinations are presented in Table D4.

Table D4. Commonality Analysis Coefficients for STAR

Variable	Coefficient	Percent For Each Variable*
Unique to Member Attendance	0.046	33%
Unique to Pre-Test Achievement Group	0.041	30%
Unique to Age range	0.020	14%
Unique to Student Recreational Attitudes	0.009	7%
Unique to 3rd Grade Proficiency Rate	0.003	2%
Common to Member Attendance, Student Recreational Attitudes	0.010	8%
Common to Member Attendance, 3rd Grade Proficiency Rate	0.007	5%
Common to Member Attendance, Student Recreational Attitudes, 3rd Grade Proficiency Rate	0.008	5%
Common to Pre-Test Achievement Group, Age Range	-0.005	-4%

* Totals will not equal 100% as they do not reflect all percentages determined by the commonality analysis.

As seen in the table above, the first three unique factors were the most powerful predictors in the model. The most practically significant predictor that can potentially be directly addressed by program staff is that of member attendance. This variable includes program administrator evaluations of a member's attendance at important events and trainings. While this variable is potentially based on intrinsic characteristics of the members, the program may be improved through a screening of potential members to maximize the rating. Interestingly, the other practically meaningful combinations in the model frequently included member attendance, highlighting the potential benefit of trying to maximize these ratings among members.

While student recreational attitudes exhibited some relative importance, the minimal effect it had uniquely and in combination with other variables may de-prioritize focus on this, especially given the potentially parent-based impact of this student attitude. Student recreational attitudes toward reading are described in more detail earlier in this section.