WestEd
WestEd.org

How Listening Drives Improvement in Vocabulary and Reading Comprehension

A Study of Promise using Tales2go

Kylie Flynn, Ph.D.
Bryan Matlen, Ph.D.
Sara Atienza, M.S.
Steven Schneider, Ph.D.

03.15.2016

Contract # 0002813.0

WestEd — a nonpartisan, nonprofit research, development, and service agency — works with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont and Georgia, to Illinois, Arizona and California, with headquarters in San Francisco. For more information about WestEd, visit WestEd.org; call 415.565.3000 or, toll-free, (877) 4-WestEd; or write: WestEd / 730 Harrison Street / San Francisco, CA 94107-1242.

© 2015 WestEd. All rights reserved.

Table of Contents

Abstract 1

Background 2

Significance 2

Study of Promise: Methods 6

Participants 6

Measures 7

Study Procedures 8

Results 11

Summary 17

References 19

Appendix 22

Abstract

This study was designed to determine the effect of adding a listening component to literacy instruction, specifically the impact on student vocabulary, reading comprehension and motivation to read. WestEd was contracted to conduct a randomized controlled trial in a San Francisco Bay area's school district's after school program. Results of the study were noteworthy, indicating Tales2go is a promising literacy tool. Students using Tales2go were approximately three months ahead in reading comprehension in just ten weeks, compared to the students who did not use Tales2go. This corresponds to a 33% increase in the rate of learning against the annual expected gain in reading achievement for that period. The Treatment group outperformed the Control group across all measures, by 3.0x in reading comprehension, nearly 7.0x in 2nd grade vocabulary, and nearly 4.0x in reading motivation. While not statistically significant, due to a limited sample size, the study effects were educationally meaningful and consistently positive across all measures. The findings further suggest a greater impact on reading achievement is possible if Tales2go is used on a regular basis, both in a classroom literacy rotation and at home.

Background

Significance

Linguistically "poor" first graders have a vocabulary of 5,000 words; linguistically "rich" first graders have a vocabulary of 20,000 words (Moats, 2001).

Problem

According to data from the 2015 NAEP Reading Assessment, 64% of all U.S. fourth grade students and 79% of all low-income students are below proficient in reading. Perhaps more concerning is NAEP Reading Assessment scores have been at these levels for decades. The problem is so severe that 32 states and the District of Columbia have statutes in place to improve third grade reading proficiency. Analysis of 2009 NAEP scores conducted by the Annie E. Casey Foundation showed students not reading at or above level by the end of 3rd grade were four times more likely not to graduate from high school.

Literacy experts have long understood that children need to be good listeners in order to become good readers. Effectively, there is a link between repeated exposure to spoken sophisticated words, vocabulary acquisition and retention and proficient reading, famously spelled out in the 'word gap' study (Hart and Risley, 1995), and recently affirmed by Stanford researchers (Fernald, 2013). Hart and Risley showed that children from low-income homes entered kindergarten having heard 32 million fewer words than children from higher income homes, the latter benefitting from a quantity of spoken, sophisticated vocabulary in the form of conversation, being read to, etc. This vocabulary deficit explained an achievement gap for low-income students that extended throughout elementary school.

Proficient reading can be expressed as the product of listening comprehension and decoding skills (Gough & Tunmer, 1986). Specifically, listening comprehension is critical to reading comprehension, since listening comprehension is language ability (i.e., vocabulary and fluency) and background knowledge (Moats, 2004). Children who can sound out words based on knowledge of letter-sound relationships (i.e., decoding) may or may not 'know' a word or know how to pronounce a word, thus lack proficiency. To become proficient readers, children need to be exposed to a rich vocabulary of words, with multiple repetitions and exposures (Senechal, 1997; Leung, 1992; Daniels, 1994, 1996; Dole, Sloan, & Trathen, 1995).

Tales2go Product

Tales2go is a literacy tool for all students, which can be especially helpful to linguistically poor students and their associated risks for poor reading outcomes. Linguistically poor first graders have a vocabulary of 5,000 words, whereas linguistically rich first graders have a vocabulary of 20,000 words (Moats, 2001). Tales2go is used in both classrooms – often as part of a literacy rotation – and at home since just half of parents read to their children every day, and just 39% for non-white families (Parenting in America Study, Pew Research Center, 2015). There are a

variety of factors that support the data, including parents who cannot read to their children in English or the estimated 32 million adults in the U.S. who cannot read at all (ProLiteracy, 2014). Furthermore, reading aloud from books is the best way to help children develop word mastery and grammatical understanding, which form the basis for learning how to read. Picture books are two to three times as likely as parent-child conversations to include a word that isn't among the 5,000 most common English words (Massaro, 2015).

Tales2go is a service that streams over five thousand name-brand audio books from leading publishers to desktops, laptops and mobile devices. Titles range from *Clifford* to *The Magic School Bus* to *Anansi* stories, as well as *Diary of a Wimpy Kid*, *Ivy + Bean*, and *39 Clues*, in addition to classics such as *Of Mice and Men* narrated by Gary Sinise. The catalog spans fiction and nonfiction, pre-K through high school and includes Spanish language titles. Tales2go is used in schools and districts across the U.S., including Los Angeles USD, Broward County Schools (FL), Heartland AEA (IA) and Orange County Public Schools (FL). Tales2go has innovated on both the delivery and business model of distributing audio books to schools. A Tales2go subscription gives educators and students instant, on-demand and simultaneous access to its catalog of titles. Competing audio book alternatives were not designed for the school market and limit schools to the titles purchased and the number of copies made available at one time, typically one at a time.

Theory of Change

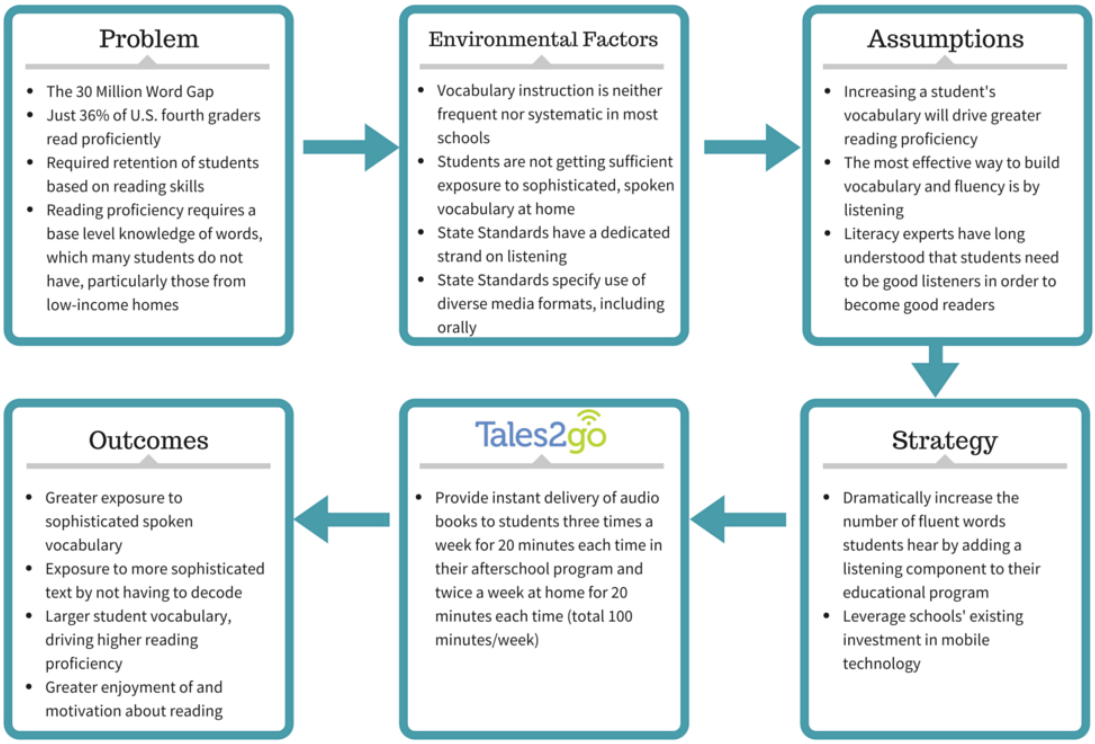

Figure 1. Theory of Change Model.

Dramatically increasing the number of fluent words a student hears is the most effective way to increase vocabulary acquisition and retention, up until late middle school, as listening comprehension outpaces reading comprehension until the age of 13 (Sticht and James, 1984). This finding extends well into middle school the work of Hart and Risley (1995) who studied

young children in their early family life and initial years at school. Not surprisingly, oral language is of the greatest importance for children most at risk – children for whom English is a second language and children who have not been exposed at home to the kind of language found in written texts (Dickinson & Smith, 1994).

Key to the link between listening and reading proficiency is that vocabulary has been empirically connected to reading comprehension since at least 1925 (Whipple, 1925) and had its importance to comprehension confirmed in recent years (National Institute of Child Health and Human Development, 2000). Yet, surprisingly, vocabulary instruction is neither frequent nor systematic in most schools (Biemiller, 2001; Durkin, 1978; Lesaux, Kieffer, Faller & Kelley, 2010; Scott and Nagy 1997). A contributing factor is likely that so much of language acquisition traditionally occurs at home, outside the classroom. Said another way, reading proficiency cannot be taught in the classroom alone. Whatever the case, too many students lack the necessary vocabulary to be successful readers, because they are not getting sufficient exposure to sophisticated, spoken words at home or in school.

Newly adopted state standards across the U.S. take this issue head-on, laying out very specific listening requirements by grade, as part of a dedicated strand on speaking and listening skills. The standards raise up listening to a skill equivalent to reading, writing and speaking, e.g., Florida's "LAFS.2.SL.1.2: Recount or describe key ideas or details from a text read aloud or information presented orally or through other media" or Virginia's "Oral Language Strand / Standard 2.2: Listen to and discuss a variety of texts."

State standards also require students to comprehend texts of steadily increasing complexity as they progress through school. The Common Core State Standards (CCSS) make the point that while reading demands in college, workforce training programs and life have generally held steady or increased in the last 50 years, K-12 texts have actually declined in sophistication, leaving a gap between ability upon high school graduation and what is expected in college and beyond. Audio books can act as an important scaffold to allow students to read above their actual grade level (Beers, 1998). And for those tempted to just pair audio with visual text, an approach which is common and valid with emerging readers, that is not what the standards intend. Listening, on its own, is considered a desired skill for a 21st-Century college and career-ready student. The CCSS requires students to read purposefully and listen attentively to gain both general knowledge and discipline-specific expertise, as well as be engaged and open-minded – but discerning – readers and listeners. Students are expected to work diligently to understand precisely what an author or speaker is saying, but also question an author or speaker's assumptions or premises and assess the veracity of claims and the soundness of reasoning. Listening is placed on equal footing with reading, writing and speaking.

Prior Research on the Benefits of Listening

Multiple studies support the premise that adding a listening component to reading instruction raises reading scores and student achievement. Among them, the Core Knowledge Foundation conducted a three-year pilot (2008-2011) with its Core Knowledge Language Arts (CKLA) program. The Core Knowledge Foundation sees reading as a complicated, cumulative process, in which both skills (e.g., decoding) and background knowledge (e.g., vocabulary and fluency) are required. In their view, decoding is a prerequisite skill, but it's not reading. To be a reader, one must have prior knowledge in order to understand the words one is being asked to decode. As a result, the CKLA program includes a listening and learning strand, which relies on read-alouds to build knowledge, vocabulary and comprehension.

The CKLA program was piloted in 10 New York City Public Schools and an additional 7 schools throughout the country, including rural and suburban schools. In total, the pilot covered 172 classrooms, 200 teachers and nearly 5,000 students of varying diversity. The percentage of students receiving free and reduced lunch ranged from 30-99%, and the percentage of students for whom English was a second language ranged from 15-60%. Results from the pilot showed CKLA students achieved significantly higher reading achievement from kindergarten to 2nd grade than a Control group of 10 schools in nearly all measures.

More recently a study was done at Cincinnati Children's Hospital, in which children were placed within an MRI machine and played pre-recorded audio books. Imaging showed that the stories read aloud activated parts of the left side of a child's brain — a region associated with understanding the meaning of words and concepts and also in memory. These same brain regions have been found to be active when older children listen to stories or read (Hutton, Horowitz-Kraus, Mendelsohn, DeWitt, Holland, and the C-MIND Authorship Consortium, 2015).

Similar Products

Audio books have been around for decades. There are several companies that distribute audio books, including Audible, OverDrive and Findaway World (maker of the Playaway). However, children's titles and the school market generally are afterthoughts for them; these players mainly focus on best-selling adult titles for the consumer and public library market. With the advent of the Internet and mobile technology, these players invested in new distribution methods, making it possible to purchase and download an audio book onto a specific device vs. selling CDs or cassette tapes. Yet despite advances in technology and delivery options, they continue to adhere to a 'lending library' business model and approach, at least with respect to the public and school library markets. In other words, a typical school library can purchase individual titles (typically $25-$60 per title) and then lend out a digital download or physical copy (e.g., a Playaway) one at a time. With those products, schools are restricted to the titles they've purchased and further restricted in the number of copies that can be checked out at one time.

Tales2go breaks new ground and stands apart from its competitors in three important ways. First, Tales2go is exclusively focused on children's titles and the school market. It is Tales2go's management's experience that focusing on a customer group leads to a better innovation for that market (e.g., Nickelodeon conceiving of a dedicated cable channel for kids vs. the broadcast networks setting aside several hours of children's programming on Saturday morning). Second, Tales2go provides instant and unlimited access to over five thousand name-brand audio books from leading publishers on desktops, laptops, and mobile devices via streaming technology. The Tales2go library is continually expanding with the addition of new publishers. Think of the difference between downloading one song on iTunes or listening to Pandora or Spotify. As with innovations in other industries (e.g., Zip cars), there is a trend away from 'ownership' towards 'access' of needed items and services. Finally, Tales2go gives educators and students simultaneous access to the titles in its library. For the first time ever, a classroom teacher can assign an audio book to an entire classroom or grade and have all the students listen at the same time on different devices. The Tales2go service is designed to accommodate the needs of a 21st Century classroom where students are issued laptops and mobile devices, or can bring their own devices into school.

Next Steps

Based on the significance of the educational problems associated with poor vocabulary and related literacy defeciencies due to students' lack of exposure to spoken, sophisticated

vocabulary, Tales2go has the potential to be an effective remediation tool. In order to avoid any bias, WestEd was contracted to conduct an independent study of promise on the impacts of Tales2go on vocabulary, reading comprehension, and motivation to read.

Study of Promise: Methods

Participants

Setting

WestEd conducted a randomized controlled trial on the use of Tales2go in a San Francisco Bay Area school district's afterschool program. The medium-sized district is economically and ethnically diverse with approximately 42% of elementary students receiving free-and-reduced lunch and 15% identified as English Learners. We worked with nine different afterschool programs at eight different school sites. The largest school had two separate afterschool programs.

Students

All 2nd and 3rd grade students enrolled in these afterschool programs were eligible for inclusion in the study provided they attended the afterschool program a minimum of three days a week. Across all sites, 127 2nd and 87 3rd graders were randomly assigned to either the Treatment condition or the Control condition (N_{treat} = 65 2nd grade students and 41 3rd grade students; N_{cntrl} = 62 2nd grade students and 46 3rd grade students; described in more detail below). All of these students received assessments in the fall prior to the implementation Tales2go intervention. See Table 1 for a breakdown of students by demographic categories.

Table 1. Demographic characteristics of participating students[1].

Female		Special education		Under-represented in STEM[2]		Socioeconomic Disadvantaged		English as Second Language	
Cntrl	Treat	Cntrl	Treat	Cntrl	Treat	Cntrl	Treat	Cntrl	Treat
45	46	7	10	52	55	55	48	21	26

[1]Five students are omitted from cell counts due to missing demographic data.

[2]Students under-represented in STEM are all non-white, non-asian students.

Measures

Reading Comprehension

The *STAR Reading Assessment* (Renaissance Learning, 2014) is a computer-based adaptive reading measure for K-12 students. It is 25 items at each grade level and takes 2nd graders under 13 minutes to complete. Extensive psychometric analyses have been conducted in the development of the STAR Reading Assessment, including item/scale calibration, reliability,

validity, and norming. Reliability scores were about 0.92, with a range 0.88 in grade 1 to 0.91 in grade 12. The STAR Reading assessment was validated using test results for more than 12,000 students from such measures as the California Achievement Test (CAT) and the Stanford Achievement Test (SAT9) and several statewide tests. The within-grade average concurrent validity coefficients for grades 1–6 ranged from 0.72–0.80, with an overall average of 0.74. Predictive validity coefficients varied from 0.69–0.72 in grades 1–6, with an average of 0.71. For the purposes of norming, analyses were conducted from a stratified random sample of approximately 70,000 students. Test scores were joined to the student-level demographics and school-level information. Sample weights from the regional, race/ethnicity, and gender results were calculated and applied to each student's ability estimate, in order to develop norms that were then transformed to the STAR Reading scaled score scale.

Vocabulary

Second Grade Vocabulary

The Vocabulary Task (VT) of the Florida Center for Reading Research (FCRR) Reading Assessment (FRA: Foorman, Petscher, and Schatschneider, 2015) was used to assess the 2nd grade students participating in the Tales2go Study of Promise. The VT is a computer adaptive task where students select the two words out of three that go together. The words are read by the computer alleviating reading as part of the task. The FRA was normed with 28,862 kindergarten to second graders throughout the state of Florida. The reliability score for the VT was .85 for second grade. Concurrent and predictive validity of .50 and .51 respectively in second grade was determined using correlations with Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007).

Third Grade Vocabulary

The Lexia Reading Assessment for Prescriptive Instructional Data (RAPID: Foorman, Petscher, and Schatschneider, 2015) was used to assess participating 3rd grade students. In the item-adaptive Vocabulary Knowledge (VK) task, the student reads a sentence that is missing a word. The student has unlimited time to select the word that best completes the sentence among three choices. The RAPID was normed using 44,780 third to twelfth grade students throughout the state of Florida. The reliability score for the VK task was 0.91. The predictive validity coefficient for the third grade VK task was .66. Construct validity was determined using the PPVT-4 (Dunn & Dunn, 2007) with an average correlation across grade levels of .52 with a range of .47 to .67. During WestEd's analyses of the data collected for the Tales2go study of promise, Ability Scores (rather than RSP or percentile scores) were used to determine the degree of growth in a skill for individual students. The Ability Score provides an estimate of a student's development in a particular skill and should change as a student's skill level increases or decreases. Each task's vertical scale has a range of 0-1000, a mean of 500, and a standard deviation of 100. Ability Scores are on an equal interval scale, so they can be added, subtracted, and used to create a mean score.

Reading Motivation

The *Motivation to Read Profile* (MRP: Gambrell, Palmer, Codling, & Mazzoni, 1996) is a reading survey developed with third and fifth graders. To determine whether the characteristics measured by the reading survey (self-concept as a reader and value of reading) corresponded to

the two subscales, factor analyses were conducted using the unweighted least squares method and a varimax rotation (Palmer et al., 1996). Only items that loaded on one of the two traits made it on the final instrument. Cronbach's alpha indicated moderately high reliability for both subscales (self-concept=.75; value =.82). Pre- and posttest reliability coefficients confirmed moderately high reliability of the instrument (self-concept=.68; value=.70).

Fidelity of Implementation

Fidelity of implementation is the degree to which an intervention or program is delivered as intended (Dusenbury, Brannigan, Falco, & Hansen, 2003). It is only through measuring the fidelity with which an intervention is implemented that a viable warrant can be made of its contribution to the desired outcomes (Carroll, et al., 2008; National Research Council, 2011, 2013). Critical components of fidelity are structural (procedural and educative) and interactional (pedagogy and student engagement; Century, Rudnick, & Freeman, 2010). In addition to a researcher designed measure of fidelity, Tales2go developers creates a teacher log that the participating afterschool program coordinators used to monitor student use of *Tales2go*. In addition to the "teacher" log, student dosage (amount of sessions and minutes) was captured via *Tales2go*'s analytics and reporting feature to determine if students use the system for 20 minutes five times a week (thrice at school and twice at home).

Study Procedures

Tales2go

WestEd researchers and employees from Tales2go conducted in-person trainings and webinars with Program Coordinators on administering assessments and implementing Tales2go, respectively, with the students during the study. The Program Coordinators were directed to have Treatment students listen to Tales2go for 20 minutes three times per week in the afternoon program at school, and to assign an additional two 20 minute sessions at home. They were also instructed to ensure that Control students did not listen or have access to the Tales2go program at any time. Tales2go employees worked with the Program Coordinators to provide them with prescribed titles and additional resources and support as needed. Tales2go also contacted parents of Treatment students regarding the home listening sessions.

Recruitment

District administrators, including the director of the after school program, provided approval for their schools to participate. Nine out of twelve after school Program Coordinators consented to participate. During the first few weeks of school, Program Coordinators sent home consent forms and project descriptions to parents/guardians of the second and third grade students in their programs. WestEd researchers assisted with recruitment by answering parent questions during student pick up, attending previously scheduled parent meetings, and offering a raffle among all consented parents. As consent forms were collected, students were excluded from participation if a Program Coordinator or parent indicated that the student would not be attending the after school program for fewer than three days per week.

Pre-assessments

WestEd researchers conducted the child assessments with some assistance from the program coordinators. All participating students were pre-assessed on the STAR Reading Comprehension test, the FRA Vocabulary Task or the RAPID Vocabulary Knowledge Task for second and third graders respectively, and the Motivation to Read Survey prior to the intervention start date of 10/19 except for three Control students who were assessed on the STAR on 10/20 and on the FRA on 10/20 and 10/21. Child assessments were administered without a fixed order. Random assignment of students to condition was conducted well after the pre-assessments were underway, so as to not influence the administration of assessments in anyway.

Fidelity

In order to capture fidelity of the intervention, all participating programs were observed by researchers two times during the 10-week intervention using a researcher-designed protocol. The fidelity protocol captured four domains of fidelity including technology, study design, and student and Program Coordinator adherence to study guidelines. The technology domain measured the extent to which the technology associated with Tales2go functioned properly. This included the Tales2go application, Chromebooks, headphones, internet access, etc. Based on the observations, the technology domain had excellent fidelity at 99% indicating the Tales2go application functioned reliably and is feasible for use in a school setting. The second fidelity domain was related to the study design and determined whether the Treatment students were receiving the intervention while Control students were not. Treatment students were expected to listen to 20 minute sessions on a dedicated device. Overall, the fidelity to the study design was good at 81% and there were no issues of contamination (i.e., Control students did not have access or exposure to Tales2go.) The duration of student listening varied. Many students listened longer than 20 minutes because they enjoyed it or they were making up for missed doses. Some students listened for less than 20 minutes if they were pulled into another activity.

The third fidelity domain related to students' adherence to the study guidelines. In particular, this domain determined if students were on task while they were listening (i.e., just listening, without paired text). Students were allowed to take notes or draw pictures while listening to the text. However, the study did not closely measure or examine whether taking notes or drawing pictures aided or impeded student attention. The focus of the study was on just listening. The overall fidelity to the study design was 62%. Researchers also recorded student actions and responses as part of this domain. Although the documentation was not comprehensive, there was a trend of favorable student responses. Some of these qualitative researcher observations included:

- Some students would express reactions while listening (e.g., laughing, gasping, etc.).
- One student asked to listen to another story, because he "liked listening".
- Another student told his Program Coordinator that he is listening at home and likes it.

The final fidelity domain related to the Program Coordinators' adherence to the study guidelines, including having a staff member monitor students while listening, completing teacher logs to track doses, and communicating with parents about the home dosage. Overall, the fidelity of this domain was good at 80% with most of the Program Coordinators exhibiting completion of the teacher logs and/or communication with parents. Researchers also recorded some comments from Program Coordinators as part of this domain with a trend of positive responses. Some of those researcher notes included:

- [Program Coordinator] said he is really happy that his students are in the program and he is making sure the children listen at school and at home.
- [Program Coordinator] said she plans to keep using Tales2go after the study and will include the students from other grades.
- I heard the Program Coordinator remind the students they should also listen for 20 minutes at home too.

> *One program coordinator stated that he is really happy that his students are receiving Tales2go and that he is making sure they are listening both at school and at home.*

The study's overall fidelity to implementation was good, particularly given the Program Coordinators' other responsibilities and competing enrichment activities in which students could participate. These factors, in part, explain reduced dosage for the majority of the sample. It is worth noting that 74% of students listened to titles on Tales2go that were beyond the study design, suggesting a broader interest in listening to audio books.

Dosage

Students in the Treatment condition were expected to use Tales2go for approximately 1,000 minutes, or 100 minutes per week, across the course of the study. Students on average used Tales2go for 604.1 minutes (standard deviation = 459.85). The observed average usage amount is approximately 60% (3/5ths) of the target usage amount. 24.5% (26 students) in the Treatment condition reached the targeted amount of listening (100 minutes per week,) noting one site completed the listening sessions in 8 weeks.

Random Assignment

The study consisted of a student-randomized control trial. Students were randomly assigned with equal chance of being assigned to either the Treatment (Tales2go) condition or the Control condition. In second grade, 65 and 62 students were randomized to the Treatment and Control conditions respectively. In third grade, 41 and 46 students were randomized to the Treatment and Control conditions respectively. Randomization occurred at the student-level by site and by grade, as such, there were Treatment and Control students within the same after school program. Issues related to potential contamination (i.e., Control students obtaining access to the Tales2go Treatment) were addressed on several levels. Tales2go developers addressed the issues of contamination directly in discussions with district-level administration and teachers. WestEd researchers addressed contamination in trainings with program coordinators and monitored it carefully during implementation. Fidelity observations reported 100% adherence to the study design, suggesting there was no evidence of Control students listening to Tales2go.

Research Questions

In order to determine the promise of Tales2go for use in a school setting, we posed the following research questions.

1. After controlling for baseline scores and student demographics, what is the impact of using Tales2go for 10 weeks on students' vocabulary, reading comprehension, and motivation to read?

2. Are these impacts moderated by grade, student demographics, and/or students' baseline performance in vocabulary, reading comprehension, or motivation to read?

Results

Attrition

Attrition in this study was defined as the percentage of students that had missing test data - either at the beginning or end of the study - as well as missing demographic information.

Two types of attrition were calculated: 1) *overall attrition* – which refers to the percentage of missing post-test data for a given outcome measure for the study sample overall (i.e., regardless of condition), and 2) *differential attrition* – which refers to the percentage of missing post-test data in the Treatment condition relative to the percentage of missing post-test data in the Control condition for a given outcome measure.

Table 2 below shows the percentages of overall and differential attrition for each outcome measure (Reading Comprehension, Grade 2 Vocabulary, Grade 3 Vocabulary, and Reading Motivation).

Table 2. Percentage of overall and differential attrition for each outcome measure.

	Overall Attrition	Differential Attrition
Reading Comprehension	17	11.5
Grade 2 Vocab	18	13.3
Grade 3 Vocab	24	21.4
Reading Motivation	17	6.9

Differential attrition values are all positive, indicating higher attrition in the Treatment relative to the Control condition. Higher attrition in the Treatment condition may have resulted from families who opted out after signing up for enrichment programs that limited their time for listening to Tales2go during the after school program. In other words, some families prioritized their student's enrichment activities over participating in the Tales2go study. As Control students did not do any Tales2go listening, their time in enrichment programs did not impact their participation in the study.

Descriptive Statistics

Prior to conducting the primary impact analyses, we examined the raw scores for students in the Treatment and Control conditions at the beginning and end of the study for all outcome measures (see Table 3). The means suggest a trend of greater growth for the Treatment group.

Table 3. Unadjusted means and standard deviations for pre-test, post-test, and growth for each outcome measure.

	reading comprehension					
	meanpre	sepre	meanpost	sepost	meangrowth	segrowth
Control	275.81	15.47	305.00	18.26	**29.19**	09.85
Treatment	314.57	22.12	357.52	25.52	**42.95**	14.53
	vocab (gr 2)					
	meanpre	sepre	meanpost	sepost	meangrowth	segrowth
Control	0.01	0.24	0.12	0.23	**0.11**	0.18
Treatment	0.47	0.21	0.76	0.25	**0.29**	0.20
	vocab (gr 3)					
	meanpre	sepre	meanpost	sepost	meangrowth	segrowth
Control	418.28	15.12	400.40	16.58	**-17.88**	15.73
Treatment	408.89	17.48	420.75	23.09	**11.86**	15.89
	reading motivation					
	meanpre	sepre	meanpost	sepost	meangrowth	segrowth
Control	44.95	1.02	44.51	0.98	**-0.44**	0.99
Treatment	42.35	1.03	43.60	1.28	**1.25**	1.06

Note: meanpre = mean score at pretest, sepre = standard error at pretest, meanpost = mean score at posttest, sepost = standard error at posttest, meangrowth = mean growth from pretest to posttest, segrowth = standard error of growth from pretest to posttest

In the next section, we examine the impact of the Treatment condition after statistically adjusting for other variables related to students' demographic and baseline achievement scores.

Impact Analyses

The goal of the impact analysis was to answer the following question: After controlling for student characteristics, what is the effect of the Tales2go intervention on students' end of study scores after a ten-week intervention period? The impact analysis specifically attempts to adjust for students' demographic information (e.g., socioeconomic disadvantaged status, special education status, gender, etc.) and test scores at the beginning of the study.

Multivariate regression was used to estimate the impact of the Tales2go intervention. The outcome variable of each measure consisted of the growth in pre- to post-test scores (calculated by subtracting the pre-score from the post-score). Each outcome growth was regressed onto the condition variable (Treatment or Control) and a host of other student characteristics, including

students' gender, special education status, socioeconomic disadvantaged status, English as second language status, STEM underrepresented status, the pre-test variable for each outcome measure, and when appropriate, students' grade, reading comprehension pre-test score, and reading motivation pre-test score.

Below are summaries of the covariate-adjusted means for each outcome measure (Figure 2) and effect sizes for each outcome (Figure 3). Figure 2 shows improvement from pre to post after Controlling for other student variables (hence adjusted means) for each condition on each outcome measure. Greater than 0 scores mean students improved on average from beginning to the end of the study. The growth for Control vs. Treatment groups was 5.70 vs 23.30 in Reading Comprehension, .03 vs .25 in Vocab Grade 2, -7.25 vs -4.90 in Vocab Grade 3, and 0.28 vs 1.36 in Reading Motivation.

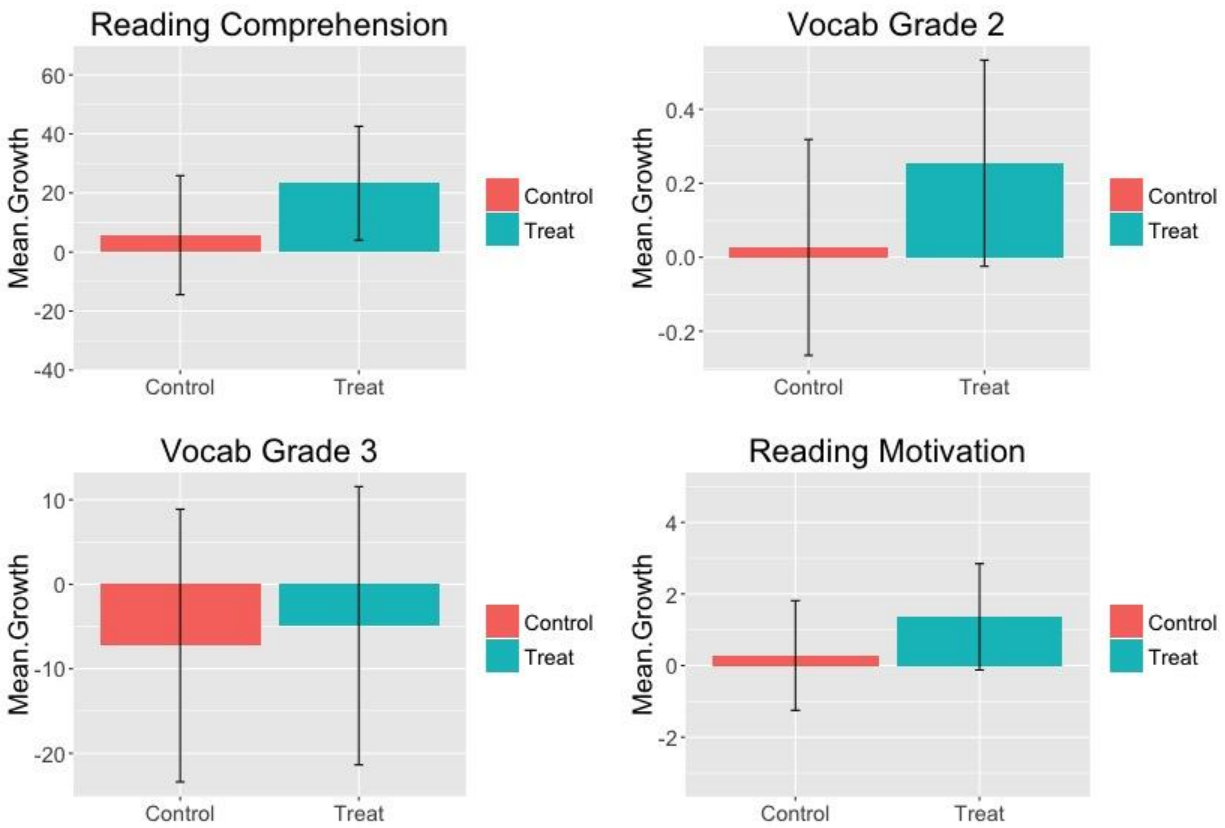

Figure 2. Covariate-adjusted mean growth and standard errors, for each outcome measure.

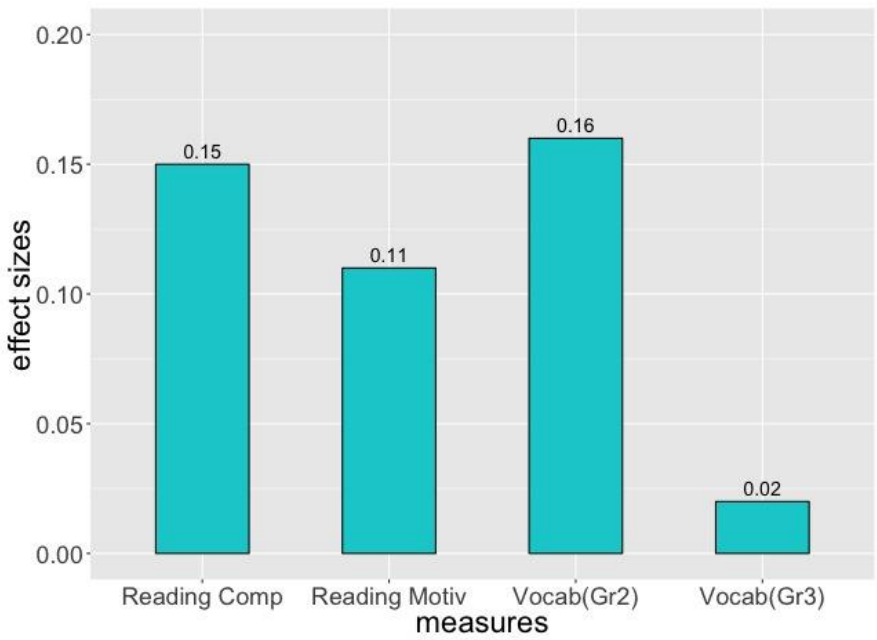

Figure 3. Effect sizes all outcome measures.

The Treatment group outperformed the Control group across all measures. While not statistically significant (all *p*s > .30; see Appendix for model summaries), the study effects were educationally meaningful and consistently positive across all measures for both 2nd and 3rd graders. In Reading Comprehension, the difference between the Treatment and Control group's growth is 17.6 (23.3 – 5.7 = 17.6), which is approximately 3 times the growth of the Control group (17.6 / 5.7 = 3.09). Using this same computation, the differential growth is approximately 7 times the growth of the Control group for Vocabulary Grade 2, 0.32 times the growth of the Control group for Vocabulary Grade 3, and 3.9 times the growth of the Control group for Reading Motivation.

Figure X shows the effect size difference between the Treatment and Control conditions for each outcome measure – these effect sizes can be interpreted as the difference (in standard deviation units) in the Treatment condition's average growth relative to the Control condition's average growth.

All effect sizes trended in a positive direction, suggesting that students in the Treatment condition exhibited higher growth than students in the Control condition. While these effect sizes are considered small, they are meaningful in the context of the effect sizes that are often observed in educational studies on elementary school children when using narrowly focused (Vocabulary) and broadly focused (Reading Comprehension) standardized tests (Hill, Bloom, Black, & Lipsey, 2008).

To interpret the effect sizes in a more practically significant way, one might convert the effect sizes based on the properties of the normal distribution. In a normal distribution, a 1 standard deviation increase corresponds to a 34 point percentile increase. Thus, in a normed sample, an effect size of .15 corresponds to a 5 point percentile increase (i.e., 34 * .15 = 5.1). Therefore, if the Control group students were at the 50th percentile of the normed sample, the Treatment group would be placed at the 55th percentile of the normed sample (i.e., 50 + 5.1 = 55.1). Effect sizes of this magnitude are also noteworthy considering the relatively short duration of the ten-week intervention.

Analyses by Grade

We also conducted the separate analyses for the Reading Comprehension and Reading Motivation measures within Grade 2 and Grade 3 students. See Figures 4 and 5, below.

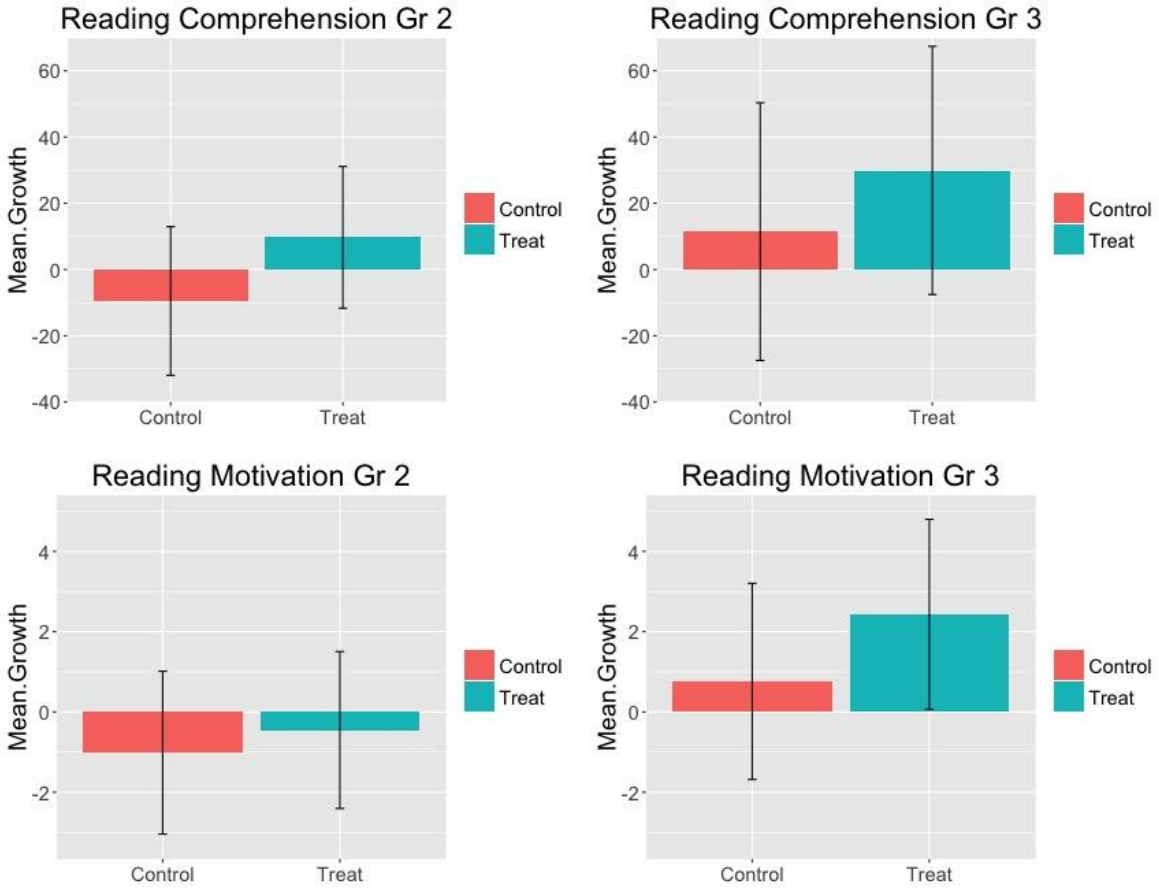

Figure 4. Covariate-adjusted means and standard errors for the Reading Comprehension and Reading Motivation measures, by condition and grade level.

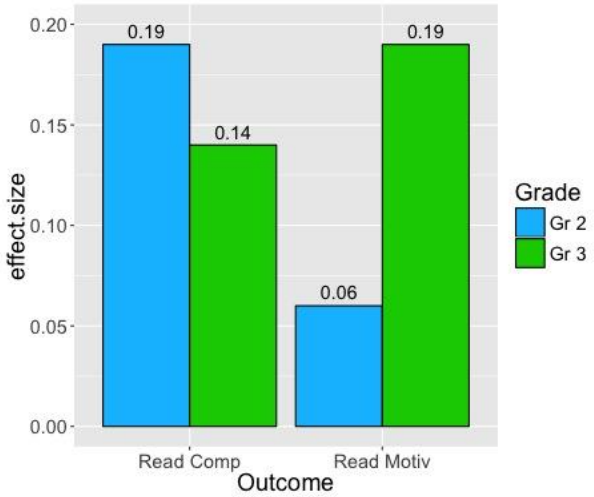

Figure 5. Effect sizes for Reading Comprehension and Reading Motivation measures, by grade.

The trends from the above analyses suggest that the impact of the Tales2go intervention consistently benefits students in the Treatment condition relative to students in the Control condition. While not statistically significant (all *p*s > .34), the study effects were educationally meaningful and positive across all measures for both 2nd and 3rd graders. In addition, the positive trends are consistent across multiple assessments covering a wide variety of constructs (e.g., reading comprehension to reading motivation).

To provide more practical significance to the effect sizes for Reading Comprehension in particular, one can compare the observed effect sizes to normative expectations for reading achievement growth. For instance, based on observed gains from nationally normed reading achievement tests, Hill, Bloom and colleagues found that the average annual gain in reading achievement for Grade 2 to Grade 3 is .60, whereas the average annual gain in reading achievement for Grade 3 to Grade 4 is .40. The effect sizes observed in the present study then correspond to 32% and 35% of the expected annual gain for Grade 2 and Grade 3 students, respectively (i.e., 19/60 = 32 and 14/40 = 35). If these effects are to be interpreted, they suggest that students using Tales2go were approximately three months ahead in reading comprehension, in just ten weeks, compared to the students who did not use Tales2go (i.e., .32 * 9 school months (i.e., the length of a school year) = 2.88 months; .35 * 9 months = 3.15 months). This corresponds to a 33% increase in the rate of learning against the annual expected gain in reading achievement for that period. It is also reasonable to assume that greater impacts would also be found in vocabulary in a study conducted over a longer period of time.

Students using Tales2go were approximately three months ahead in reading comprehension in just ten weeks. This corresponds to a 33% increase in the rate of learning against the annual expected gain in reading achievement for that period.

Moderator Analyses

We next conducted moderator analyses. A moderator effect refers to when the effect of the Treatment depends on a level of one of the covariates. For example, perhaps the intervention is more effective for girls than for boys. To explore this possibility, we reran the models described above including interaction terms for each covariate with the condition variable (i.e., covariates consisted of students' gender, special education status, socioeconomic disadvantaged status, English as second language status, STEM underrepresented status, the pre-test variable for each outcome measure, and when appropriate, students' grade, reading comprehension pre-test score, and reading motivation pre-test score). Because the sample in Grade 3 students was small to begin with, we ran this analysis only in the Grade 2 students.

For grade 2 students, the only consistent moderating variable was the baseline Reading Motivation scores. The Reading Motivation scores at baseline moderated the effect of the Treatment (at the alpha < .10 level). The moderation effect was such that higher scores on baseline motivation (i.e., pre-test) led to higher growth in motivation ($\beta = .36$, $p = .06$) and higher growth in Vocabulary ($\beta = .05$, $p = .09$).

Summary

Tales2go shows good promise as a literacy tool to improve vocabulary, reading comprehension, and motivation to read. Overall, Treatment students showed greater growth in their vocabulary, reading comprehension, and motivation-to-read than did students in the Control group. Specifically, the effects suggest students using Tales2go were approximately three months ahead in reading comprehension, in just ten weeks, compared to the students who did not use Tales2go. This is particularly compelling given the brevity of the intervention and lower-than-expected dosage. Moreover, the positive outcomes are likely to improve with increased exposure to Tales2go. Finally, qualitatively, researcher observations and notes indicated that both educators and students liked the Tales2go program. Specifically, Program Coordinators valued having their students listening and the students appeared to enjoy their time listening.

Limitations

The study findings are even more impressive when considering the limitations of the study, which include a relatively short intervention, a limited sample size, and low dosage relative to study design. We consistently observed meaningful positive effects suggesting higher growth for students who used Tales2go. These positive effects were observed in a varied set of outcomes, including reading comprehension, vocabulary, and motivation for reading. These consistently positive findings highlight the potential impact of Tales2go when implemented in a classroom literacy rotation (i.e., consistent and increased dosage over a longer period of time).

Directions of Future Research

In addition to the duration and dosage of the intervention, there are other findings that warrant further research. In particular, the outcomes on vocabulary were more meaningful for 2nd graders than for 3rd graders. Given the importance of early literacy skills, including vocabulary development, to students' later reading ability (Whitehurst and Lonigan, 1998), it is not

surprising that younger students benefitted more from Tales2go than their older peers. The longer educational providers wait to address the critical components underlying reading development, the more difficult it becomes to make an impact. This finding suggests that it will be important to explore the impact of Tales2go with students in kindergarten and first grade as well, in order to determine the ideal age range for maximizing the impacts. Strengthening the case for implementing Tales2go with younger students is the finding that reading motivation at pretest was associated with higher growth on reading motivation and vocabulary at posttest for second graders. Thus, it is worth exploring whether or not capitalizing on the motivation of younger students creates improved outcomes for students in grades K-2 when compared to students in grades 3-5. Finally, it is worth exploring whether improved dosage can be obtained in a more structured educational setting, such as school libraries, the general education classroom, and/or resource centers for special education or ELL students, with a literacy rotation, small group guided reading, or independent reading time.

References

Beers, Kylene (1998). Listen While You Read: Struggling Readers and Audiobooks. School Library Journal 44 (4): 34-35.

Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. The American Educator, 25(1), 24-28.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. & Balain, S. (2008). A conceptual framework for implementation fidelity. Implementation Science, 2(40).

The Annie E. Casey Foundation. (2013). Early Warning Confirmed: A Research Update on Third Grade Reading. Baltimore, MD: Author. Retrieved from www.aecf.org

The Annie E. Casey Foundation. (2013). Early Warning Confirmed: A Research Update on Third Grade Reading. Baltimore, MD: Author. Retrieved from www.aecf.org

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*(2), 199-218.

Core Knowledge Foundation (2011). CKLA Three-Year Pilot Program Results. New York, NY: Author. Retrieved from http://www.coreknowledge.org/language-arts-program-pilot

Daniels, M. (1994). The effect of sign language on hearing children's language development. Communication Education, 43(4), 291-298.

Daniels, M. (1996). Seeing language: The effect over time of sign language on vocabulary development in early childhood education. Child Study Journal, 26, 193-208.

Dickinson, D.K., & Smith, M.W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly, 29,* 104-122.

Dole, J.A., Sloan, C., & Trathen, W. (1995). Teaching vocabulary within the context of literature. Journal of Reading, 38(6), 452-460.

Dun, L., & Dunn, D. (2007). *Peabody Picture Vocabulary Test-4.* San Antonio, TX: Pearson.

Durkin, D. (1978-79). What classroom observations reveal about reading comprehension instruction. Reading Research Quarterly, 14, 481-533.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W.B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18,* 237-256.

Foorman, B.R., Petscher, Y., & Schatschneider, C. (2015). *Florida Center for Reading Research (FCRR) Reading Asssessment (FRA).* Tallahassee, FL: Florida State University

Foorman, B.R., Petscher, Y., & Schatschneider, C. (2015). *Reading Assessment for Prescriptive Instructional Data (RAPID).* Concord, MA: Lexia Learning Systems

Fernald, A., Marchman, V., & Weisledger, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. Developmental Science, 16(2), 234-248.

Gambrell, L.B., Palmer, B.M., Codling, R.M., & Mazzoni, S.A. (1996). Assessing motivation to read. *The Reading Teacher*, 49, 518-533.

Gough, P.B. & Tunmer, W.E. (1986). Decoding, reading, and reading disability. Remedial and Special Education, 6, 6-10.

Hart, B., & Risley, R.T. (1995). Meaningful differences in the everyday experience of young American children. Baltimore: Paul H. Brookes

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177.Hutton, Horowitz-Kraus, Mendelsohn, DeWitt, Holland, and the C-MIND Authorship Consortium, 2015

Lesaux, N.K., Kieffer, M.J., Faller, E., & Kelley, J. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. Reading Research Quarterly, 45(2), 198-230.

Leung, C.B. (1992). Effects of word-related variables on vocabulary growth repeated read-aloud events. In C.K. Kinzer & D.J. Leu (Eds.), Literacy research, theory, and practice: Views from many perspectives: Forty-first Yearbook of the National Reading Conference (pp. 491-498). Chicago, IL: The National Reading Conference.

Massaro, D.W. (2015). Two different communication genres and implications for vocabulary development and learning to read. University of California, Santa Cruz.

Moats, L.C. (2001). When older students can't read. Educational Leadership, 58(6), 36-40.

Moats, L.C. (2004). Language essentials for teachers of reading and spelling (LETRS), Module 2, The Speech Sounds of English, and Module 3, Spellography for Teachers. Longmont, CO: Sopris West Educational Services

National Research Council, (2011). Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics. Committee on Highly Successful Science Programs for K-12 Science Education. Board on Science Education and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council, (2013). Monitoring progress toward successful K-12 STEM education: A nation advancing? Committee on the Evaluation Framework for Successful K-12 STEM Education. Board on Science Education and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Institute of Child Health and Human Development. (2000). Report on the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Pew Research Center (2015). Parenting in America Study. Retrieved from http://www.pewsocialtrends.org/2015/12/17/parenting-in-america/

ProLiteracy (2014). The Numbers Don't Lie. Retrieved from http://www.proliteracy.org/the-crisis/adult-literacy-facts

Scott, J.A., & Nagy, W.E. (1997). Understanding the definitions of unfamiliar words. *Reading Research Quarterly, 32,* 184-200.

Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. Journal of Child Language, 24(1), 123-138.

Sticht, T.G., & James, J.H. (1984). Listening and reading. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of Reading Research: Volume 1* (pp. 293-317). White Plains, NY: Longman.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 and 2015 Reading Assessment.

Whipple, G. (Ed.). (1925). The Twenty-Fourth Yearbook of the National Society for the Study of Education: Report of the National Committee on Reading. Bloomington, IL: Public School Publishing Company.

Whitehurst, G.J. & Lonigan, C.J. (1998). Child development and emergent literacy. *Child Development, 69*(3), 84.

Appendix

Baseline Equivalence

Table 4 below shows the effect size differences at baseline for the baseline (i.e., sample at the onset of the study) and analytic samples (i.e., sample after attrition on the post-tests and covariates).

Table 4. Baseline equivalence effect sizes (in Hedges *g*) for the baseline and analytic samples. Negative effect sizes indicate higher scores for the Control relative to the Treatment condition at baseline, and vice versa for positive effect sizes.

	Baseline Sample	Analytic Sample
Reading Comprehension	.07	.21
Grade 2 Vocab	.17	.27
Grade 3 Vocab	-.22	-.10
Reading Motivation	-.18	-.29

Model Summaries

Table 5. Reading Comprehension model summary

	estimate	st.error	p_value
(Intercept)	50.2	18.59	0.01
Condition	17.6	17.28	0.31
Reading Comprehension PreTest (grand mean centered)	-0.18	0.06	0
Grade	51.04	18.19	0.01
Reading Motivation Pretest (grand mean centered)	-1.26	0.9	0.16
Gender	-12.75	17.07	0.46
ESL	-24.17	21.81	0.27
Stem Underrepresented	-1.13	19.61	0.95
Special Education	-53.56	32.72	0.1
Socioeconomic Disadvantaged	-51.9	21.13	0.02

Table 6. Reading Motivation model summary

	estimate	st.error	p_value
(Intercept)	-3.72	1.42	0.01
Condition	1.08	1.32	0.42
Reading Motivation (grand mean centered)	-0.47	0.07	0
Grade	5.1	1.4	0
Reading Comprehension (grand mean centered)	-0.01	0	0.01
Gender	-0.96	1.31	0.46
ESL	-0.14	1.66	0.93
Stem Underrepresented	2.07	1.51	0.17
Special Education	0.73	2.49	0.77
Socioeconomically Disadvantaged	0.85	1.61	0.6

Table 7. Vocab Grade 2 model summary

	estimate	st.error	p_value
(Intercept)	0.58	0.25	0.02
Condition	0.23	0.25	0.36
Vocab Pretest (grand mean centered)	-0.52	0.09	0
Reading Motivation (grand mean centered)	0.01	0.01	0.62
Reading Comprehension (grand mean centered)	0	0	0
Gender	-0.1	0.25	0.68
ESL	-0.1	0.31	0.75
Stem Underrepresented	0.16	0.31	0.61
Special Education	0.07	0.47	0.89
Socioeconomically Disadvantaged	-0.8	0.34	0.02

Table 8. Vocab Grade 3 model summary

	estimate	st.error	p_value
(Intercept)	-41.64	25.05	0.1
Condition	2.34	22.71	0.92
Vocab Pre-test (grand mean centered)	-0.47	0.13	0
Reading Motivation (grande mean centered)	-1.89	1.19	0.12
Reading Comprehension (grand mean centered)	0.25	0.08	0
Gender	10.95	22.05	0.62
ESL	-22.35	30.26	0.46
Stem Underrepresented	26.6	24.64	0.28
Special Education	1.34	47.85	0.98
Socioeconomically Disadvantaged	7.45	25.77	0.77

WestEd
WestEd.org
730 Harrison Street
San Francisco, California 94107-1242